New approaches to unveil the Transcriptional landscape of dopaminergic neurons

Roberto Simone

SISSA

International School for Advanced Studies

PhD dissertation in Structural and Functional Genomics

Trieste – October 2008

Supervisor: Prof. Stefano Gustincich, Ph.D.
Neurobiology Sector
The Giovanni Armenise-Harvard Laboratory
SISSA - Trieste

External Examiner: Dr Valerio Orlando, Ph.D.
Epigenetics and Genome Reprogramming group
Dulbecco Telethon Institute
IGB CNR - Naples

"La gente di solito usa le statistiche come un ubriaco i lampioni: più per sostegno che per illuminazione".
Mark Twain

"L`insieme e` piu` della somma delle sue parti"
Aristotele

"La scienza della complessità` ci insegna che la complessita` che vediamo nel mondo e` il risultato di una semplicita` nascosta".
Chris Langton

To my Parents: Cinzia and Giovanni,
and to my sister Patrizia

Table of contents

SECTION 1. INTRODUCTION

The complexity of the Eukaryotic Transcriptome

Technologies to unveil transcriptome complexity

CAGE data and the analysis of mammalian promoters

Mesencephalic Dopaminergic Cells

Mesencephalic Dopaminergic Cells: DAT, VMAT2 and COMT are crucial components of their function

Parkinson`s disease

SECTION 2. MATERIALS AND METHODS

SECTION 3. RESULTS

Developing of a new tecnique for Genome-Wide Tagging and Mapping of the Transcription Start Sites from limited amount of RNA: nanoCAGE and high-throughput sequencing

Application of the NanoCAGE technology to the mesencephalic dopaminergic neurons SN (A9) and VTA (A10) dopaminergic nuclei

Use of alternative promoters of biologically relevant protein coding genes in SN and VTA neurons

The non coding RNAs landscape in SN and VTA neurons as emerging from nanoCAGE and deep sequencing

SECTION 5. DISCUSSION and CONCLUSION

Acknowledgements

References

Abstract

Recent advances in studying the mammalian transcriptome arised new questions about how genes are organized and what is the function of noncoding RNAs. Furthermore, the discovery of large amounts of polyA⁻ transcripts and antisense transcription proved that a portion of the transcriptome has still to be characterized.
The complex anatomo-functional organization of the brain has prevented a comprehensive analysis of the transcriptional landscape of this tissue. New techniques must be developed to approach neuronal heterogeneity.

In this study we combined Laser Capture Microdissection (LCM) and nanoCAGE, based on Cap Analysis of Gene Expression (CAGE), to describe expressed genes and map their transcription start sites (TSS) in two specific populations, A9 and A10, of mouse mesencephalic dopaminergic cells. Although sharing common dopaminergic marker genes, these two populations are part of different midbrain anatomical structures, substantia nigra (SN) for A9 and ventral tegmental area (VTA) for A10, project to relatively distinct areas, participate to distinct ascending dopaminergic pathways, exhibit different electrophysiological properties and different susceptibility to neurodegeneration in Parkinson`s disease.
Specific neurons were identified by the expression of Green Fluorescent Protein driven by a cell-type specific promoter in transgenic mice. High-quality RNAs were purified from 1000-2500 cells collected by LCM.

We adapted the CAGE technique to analyze limiting amounts of RNAs (nanoCAGE). We took advantage of the cap-switching properties of the reverse transcriptase to specifically tag the 5`end of transcripts with a sequence containing a class III restriction site for EcoP15I. By creating 32bp 5`tags, we considerably improved the TSS mapping rate on the genome. A semi-suppressive PCR strategy was used to prevent primer dimers formation. The use of random priming in the 1st strand synthesis allowed to capture poly(A)⁻ RNAs. 5`tags were sequenced with Illumina-Solexa platform.

Here we show that this new nanoCAGE technology ensures a true high-throughput coverage of the transcriptome of a small number of identified neurons and can be used as an effective mean for gene discovery in the noncoding RNAs, to uncover putative alternative promoters associated to variants of protein coding transcripts and to detect potentially regulatory antisense transcripts.

A further experimental validation by 5`RACE (Rapid Amplification of cDNA Ends) and RT-PCR on few candidate genes, have confirmed the existence *in vivo* of alternative TSS in the case of key regulatory genes involved in specifying and maintaining the dopaminergic phenotype of these neurons such as α-synuclein (Snca), dopamine transporter (Dat), vescicular monoamine transporter 2 (Vmat2), catechol-O-methyltransferase (Comt). Furthermore the differential expression of an antisense transcript overlapping to the polyubiquitin (Ubc) gene was detected as potentially interesting candidate gene accounting for differences in the ubiquitin-proteasome system (UPS) function in the two neuron populations. The potential implications deriving from these newly discovered alternative promoters and transcripts are discussed, considering also the potential consequences for the corresponding protein isoforms.

List of abbreviations

α-syn	α-synuclein
A10	mesencephalic dopaminergic neurons of the VTA
A9	mesencephalic dopaminergic neurons of the SN
ACh	acetylcholine
AD	Alzheimer`s disease
AS	antisense transcript
CAGE	Cap Analysis of Gene Expression
CNS	central nervous system
COMT	catechol-O-methyl transferase
CPu	caudate-putamen
DA	dopamine
DAT	dopamine transporter
DBH	dopamine-β-hydroxylase
DLB	Dementia with Lewy Bodies
DOPA	3,4-dihydroxyphenylalanine
DOPAC	3,4-dihydroxyphenylacetic acid
ENCODE	ENCyclopedia Of DNA Elements project
FANTOM	Functional Annotation of Mouse project
GABA	γ-aminobutyric acid
Glu	Glutamate
5-HT	5-hydroxytryptamine, serotonin
HVA	homovanillic acid
IVT	*In vitro* transcription
LCM	Laser Capture Microdissection
LTD	Long Term Depression
LTP	Long Term Potentiation
MAO	monoamine oxidase
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
mRNA	messenger ribonucleic acid
MSA	Multiple System Atrophy
3-MT	3-methoxytyramine
NAc	nucleus accumbens
NAC	non-amyloid β component of Alzheimer`s senile plaques
NACP	non-amyloid β component containing protein
ncRNA	non protein coding transcript
NMDA	N-methyl-D-aspartate
6-OHDA	6-hydroxydopamine
PC	Parametric clustering
PD	Parkinson`s disease
PFC	Prefrontal Cortex
RACE	Rapid Amplification of cDNA ends
ROS	Reactive Oxygen Species
RT-PCR	Reverse Transcription - Polymerase Chain Reaction
S	sense transcript
SAGE	Serial Analyis of Gene Expression
SMART	Switching Mechanism At the 5`end of RNA Transcripts
SN	substantia nigra
SNc	substantia nigra pars compacta
SNr	substantia nigra pars reticolata
TC	Tag proximity clustering
TSS	Transcription Starting Site
TUF	Transcript of unknown function
UTR	untranslated region
VMAT2	vescicular monoamine transporter 2
VTA	ventral tegmental area

SECTION 1. INTRODUCTION

The complexity of the Eukaryotic Transcriptome: Large part of the genome is transcribed

In recent years a lot of efforts were made to determine the genomic sequence of most of the so called "model organisms" in order to identify how many are the genes and to clarify which is their structure.

With the recent availability of genomic sequences, publicly accessible in databases, it became evident that the number of transcripts exceeds of at least one order of magnitude the number of annotated genes along the chromosomes.

Many independent experimental works in the last five years, using different technologies to explore genome-wide gene expression, conveyed to a common surprising conclusion: the genome is much more active than previously thought as transcription is widespread than existing genome annotations would predict.

And this is clearly contraddicting what only few years ago was regarded as the functional interpretation of the genome.   In fact, according to the classically accepted view, about 97% of the genome was defined as "junk DNA" due to the lack of a clear associated function whereas only 3% was associated with protein-coding sequences, and "genes" were generally assumed to be synonymous of "proteins", or protein-coding sequences. In the past decades few pioneering studies showed evidences of a much larger complexity of the transcriptomes . In the 1970s, in sea hurchin embryos heterogeneous nuclear RNA (hnRNA) were found exibihiting 10-fold more complexity than the cytoplasmic fraction of mRNA associated with polysomes. In the 80s it was shown that in amphibian oogenesis the transcription levels are an order of magnitude greater than expected for mRNAs alone (Varley et al. 1980). Furthermore, an evaluation based on nucleic acid reassociation kinetics found 10-fold more hnRNA than mRNAs (Hough et al. 1975). In human cells the nuclear RNA fraction showed to be 10-fold more complex than the cytoplasmic fraction and 5'capped

RNAs outnumbered poly-adenilated RNAs 3:1 in hamster ovary cells (Salditt-Georgieff et al. 1981). All of these cases were considered with skepticism and treated more like exceptions than like a general property of genomes.

More recently, some important technological advancements made possible to take a global snapshot of the output of the genome, shading some lights on the nature of this transcriptional diversity. Using genomic tiling arrays, a widespread transcription along most of the human chromosomes has been observed including a significant proportion of antisense transcription as well as many non coding and intergenic transcripts (Kapranov et al. 2002; Bertone et al. 2004; Cheng et al. 2005). These data were largely confirmed and expanded by studies involving other technologies like full-length cDNA sequencing (Okazaki et al. 2002), SAGE tags expression profiles (Chen et al. 2002; Saha et al. 2002), and CAGE tags profiling (Carninci et al. 2005; Carninci et al. 2006). All of these evidences pointed out to disclose a hidden layer of transcriptional acitivity mostly arising from non coding regions, intergenic trans-acting regulatory regions, mobile genomic elements, pseudogenes. All these were elements previously regarded as a silent part of the genomic structures without any associated function.

Clearly, one cannot completely rule out that transcription in mammals could be inherently sloppy and at least part of those non coding transcripts may represent an unexplained large excess of "unspecific transcription". However, many of them are expressed in a developmentally regulated manner, being gender-, or tissue- or cell-specific, which suggest that most of them are not spurious. So it may be considered true, for instance, that the number of human genes, considered as tracts of genomic DNA transcribed to produce functional information, is much higher than previously anticipated solving at least in part the diatribe between evaluations based on genomic sequences analysis (30-40,000) (Lander et al. 2001; Venter et al. 2001) or cDNA cluster analysis (65-70,000) (Zhuo et al. 2001).

However, although in the past few years tremendous efforts were made to assembly the complete sequence of the genome of a growing number of different organisms, we are still far away to have a

comprehensive view of all the molecular diversity of the corresponding transcriptomes. This is mainly due to the lack of a uniform methodological approach and of a systematical collection of samples. Since the transcriptome discovery rate has not reached a plateu, this is a still growing process (Carninci P. 2007).

The complexity of the Eukaryotic Transcriptome: alternative splicing, alternative transcription initiation and termination

After the discovery of exons and introns in the Adenovirus *hexon* gene in 1977 (Sambrook J. 1977) Walter Gilbert proposed that different combinations of exons could be spliced together to produce different mRNA isoforms of a gene (Gilbert W.1978). In the `80s researchers documented alternative splicing in several genes and estimated that 5% of the genes in higher eukaryotes might have alternative splicing (Sharp P.A. 1994). Now we know this represents a much more common mechanism to create transcript diversity.

Most of the eukaryotic genes exhibit a completely different structure from Prokaryotes, being composed from small coding sequence tracts (exons) interconnected by large non coding segments known as introns. The joining of the coding portions in different orders, a process known as "alternative splicing", may originate additional transcripts diversity in the respective protein products. This diversity can explain, at least in part, how the relatively similar numbers of protein-coding genes estimated for fruit fly (13,985 BDGP release 4), nematode C.elegans (21,009 Wormbase release 150), and human (23,341 NCBI release 36) result in the remarkable phenotypic differences observed among these species.

Exon inclusion into a mature mRNA is a complex choice whose outcome is influenced by a variety of different factors. Many exons are "cryptic", in the sense that they are included only in some but not all transcripts, through mechanisms that are so far not completely understood. On average, cryptic exons are reported to be shorter than constitutive exons (Berget 1995) and to be flanked by

weaker splice sites (Stamm et al. 2000). Although is known that exonic splice enhancers can compensate for weaker splice sites (Berget 1995, Fairbrother et al. 2002), it is unclear at present if cryptic and constitutive exons differ in the content of splice enhancers.

The percentage of intronic genes vary among species, as well as the intron : exon ratio, but in mammals the vast majority of these genes are intronic.

The developing of better strategies for full-length cDNA cloning together with the advent of genomic tiling arrays and an improvement in the algorithms for gene structure prediction revealed the existence of an hidden layer of complexity even in the protein-coding genes. The emerging view is that most of the "transcriptional units" present, at different extent, alternative splicing, alternative transcription initiation and termination.

In mouse, the completion of the genomic sequence (ftp://wolfram.wi.mit.edu/pub/mousecontigs/MGSCV3) and the availability of high quality full-length cDNA libraries obtained with the cap trapping method on polyadenilated and capped RNAs (Carninci et al.1996) made possible to analyze large datasets of multiple full length cDNA sequences derived from the same "transcription unit", revealing the entire spectrum of exon combinations and contributing to disclose the *cis*-acting sequences which mediate the choice of each particular exon. In a recent study, an extensive analysis of the data coming from RIKEN Fantom 2 mouse 60,770 full-length cDNA collection and 44,122 public mRNA sequences, revealed that at least 41% of the cDNAs exihibit at least one variant form. 54,490 (70%) out of 77,640 cDNA sequences mapping at 95% identity on the mouse genome were "multiexon transcripts". 7,293 of these multiexon transcripts (13%) didn`t cluster with any transcripts and they were called singletons whereas the remaining part (47,197 cDNA sequences) were multiexon transcripts clustering into 11,677 multitranscript clusters. Of these, 41% were variant, originated by alternative splicing events (Zavolan et al. 2003).

However, the real value of alternative spiced forms could be much higher as predicted from the fact that the estimated frequency of genes with alternative splicing increases with the depth of sampling

and with the average number of transcripts sequenced for a gene (Zavolan et al 2002; Kan et al. 2002). In the same study previously reported, a Bayesan statistical model was constructed to estimate the probability that a gene has multiple splice forms even though in the case when only an identically spliced form was observed in the experimental dataset. This mathematical model set the upper bound for the incidence of genes with multiple splice forms (Zavolan et al. 2003).

This estimation fits well also with other studies in human where between 40% and 60% of all the genes were predicted to have alternative splice forms (Modrek et al. 2001, Modrek B. and Lee C. 2002). This is mainly based on EST and available mRNA sequences alignment to the genomic sequence. More recently also tiling array analysis of human chromosome 21 and 22 showed that the number of detectable transcribed exons that are expressed in at least one out of 11 different cell lines tested is 10-fold larger than the number of the currently annotated exons (Kampa et al. 2004). All these studies show that the alternative splicing is far more abundant, ubiquitous and functionally relevant than previously thought.

Despite an initial study where most alternative splicing events occur within the 5` untranslated regions (Mironov et al. 1999), more recent reports indicate that 70-88% of the alternative splices change the protein products, having functional impact such as replacement of the amino or carboxy terminus, or in-frame addition and removal of a functional domain (Modrek B. and Lee C. 2002). Furthermore, the alternative splicing events tend to preserve the integrity of the coding frame since in human only 19% of the alternative protein forms are shortened due to frameshifts (Modrek et al. 2001). The functional pattern of alternative splicing across the genome seems to affect mainly certain categories of genes as in a random sample of 50 human genes, over 75% were involved in signalling, regulation and transcription, with a strong occurrence for genes expressed in immune and nervous systems (Modrek B. and Lee C. 2002). A very low occurrence frequency is observed for genes coding for enzymatic proteins implying that may be a very limited potential for diversification of enzymatic function through splice variation (Zavolan et al. 2003).

In general multiple alternative splicing in the same transcript is not so frequent, However, notable examples of combinatorial alternative splicing of multiple exons, generating up to 40,000 isoforms of a single gene, has been recently discovered in the nervous system, including *Dscam* (axonal guidance receptor in *Drosophila*) and neurexin, a neuropeptide receptor (Graveley B.R. 2001). Some additional interesting considerations about the role played by alternative splicing in the gene function regulation come from the analysis of CAGE tag libraries.

Mapping the transcription start sites (TSS) with CAGE technique it was shown that in mouse there are at least 236,000 TSS and 153,000 transcription termination sites (TTS) for only 44,000 transcriptional units, implying that many transcriptional units have large TSS-covered regions and TTS variability. On average, a defined 3'end has 1.32 start sites while for every 5'end an average of 1.83 of 3'ends are associated (Carninci et al. 2005). These data are also confirmed by a comprehensive long-SAGE analysis in mouse with associated 3.3 alternative 3'ends to each transcript (Siddiqui et al. 2005).

Alternative TSS usage could be important for driving the expression of different mRNA isoforms. Clustering tags-derived TSS based on their CAGE-determined level of expression showed that different TSSs belonging to the same transcription unit fall into different expression patterns.

Furthermore, 58% of protein-coding genes had two or more alternative promoters based on non-overlapping tag clusters. The variability at the TTS could have a relevant role for mRNA localization and stabilization as suggested by the fact that 27 sequence motifs are statistically over-represented within 120bp of the polyadenilation site (Carninci et al. 2006).

The dark side of the Eukaryotic Transcriptome –Transcripts of Unknown Function (TUF)

A large fraction of the genome of higher eukaryotes does not encode for proteins at all (Eddy 1999; Erdmann et al. 1999; Mattick and Gagen 2001). The initial discovery of many non coding transcripts was limited to some isolated cases. However, independent experimental studies reported

the existence of a growing number of ncRNAs. Well documented examples comprise XIST, involved in female X chromosome inactivation (Brockdorff 1998; Hong et al. 2000), and the H19/*Igf2* locus, whose mutation affect cell proliferation (Wrana 1994; Hurst and Smith 1999) both of which are imprinted and differentially spliced without encoding any protein. Other examples are *roX1* and *roX2* RNAs involved in male X-chromosome activation for the dosage response in *Drosophila*, heat shock response RNAs in *Drosophila*, oxidative stress response RNAs in mammals, His-1 RNA involved in viral response/carcinogenesis in humans and mice, SCA8RNA which is antisense to an actin-binding protein mRNA and mutated in spinocerebellar ataxia type 8 (Eddy 1999; Erdmann et al. 1999; Prasanth and Spector 2007).

Many examples were found in imprinted loci. In the murine *Igf2r/Air* locus, a differentially methylated region (DMR2) within the promoter of the ncRNA *Air* act as a critical bidirectional element, controlling silencing of the paternal allele of three protein-coding genes Igf2r, Slc22a2 and Slc22a3 (Zwart et al. 2001). Another locus (*KCNQ1*) is associated with human Beckwith-Wiedemann syndrome (BWS) whose etiology is complex because involves many genes at two different loci, and where the transcription of a paternally expressed ncRNA appears to be critic to establish the imprinted profile of the nearby genes (Szymanski and Barciszewski 2003; Mancini-Dinardo et al. 2006). Interestingly, the ncRNA promoter lies within a differentially methylated region overlapping with the silenced gene (Beatty et al 2006).

One of the first examples of imprinted disorder to uncover the role of ncRNA genes was the Prader-Willi/Angelman syndrome (PWS/AS). This results from the distrupted expression of imprinted genes spanning a region of more than 4Mb on human chromosome 15 (mouse proximal 7). Maternal disomies result in PWS, whereas paternal disomies result in AS (O`Neill 2005). In the case of PWS the expression of a group of paternally transcribed protein genes (*SNURF/SNRPN*, *MKRN3*, *MAGEL2*, *ZNF127*) is lost and a spliced and polyadenylated ncRNA (*IPW*) has been suggested to control silencing and expression at this chromosomal region. At the same locus the

ncRNA (*ZNF127AS*) is transcribed antisense to a brain and lung expressed transcript (Wevrick et al. 1994; Jong et al. 1999).

AS disorder is characterized by loss of function mutations in a maternally transcribed locus *UBE3A*. The paternal silencing of *UBE3A* locus is confined to brain subregions and interestingly there is a paternal-specific expression of a large, alternatively spliced antisense transcript (*UBE3ATS*) spanning for about 450 kb in human and 1Mb in mouse, whose expression may be directly linked to the etiology of the disease (Runte et al. 2004). The PWS/AS locus also contains several clusters of mostly intronic C/D-box snoRNAs, exclusively expressed from the paternal chromosome. Most of these snoRNAs, which overlap the *UBE3A* gene on the opposite strand, are overexpressed in AS patients (Runte et al. 2001).

The intergenic transcription has been observed to produce many ncRNAs which can play essential roles in the coordination of gene expression, considering the fact that most of the regulatory elements of the genome like enhancers, insulators, genomic bordering elements and locus control regions (LCR) are often located outside the coding tracts of the protein-coding genes. The active transcription of these intergenic regions seems to suggest that they can exert their effects, at least in part, in a *trans*-acting way (Mattick and Gagen 2001).

In many cases non coding genes are expressed in a developmentally-regulated manner like in *Drosophila* for the Bithorax/Abdominal A-B locus. This comprises at least 7 major transcriptional units. While three of them encode for protein products, some have associated phenotypic signatures and are developmentally regulated (Akam et al. 1985; Lipshitz et al. 1987; Sanchez-Herrero and Akam 1989).

Another example of intergenic transcription comes from the β-globin locus, which was discovered long time ago to exhibit the constitutive production of specific ncRNAs from both intergenic regions and the locus control region (LCR) (Ashe et al. 1997). Several models have been advanced to find out a role for these non coding transcripts in the regulation of the locus. In the so called "tracking model" the recruitment of erythroid-specific and ubiquitous transcription factors may

account for helping the transcription initiation from the promoters of protein encoding genes of the β-globin locus (Li Q. et al. 2002). In a second model, intergenic transcription may be required to establish and maintain the open state of the chromatin in the locus. The persistence of DNase I hypersensitivity following deletion of the LCR in cell lines argue against this role (Gribnau et al. 2000; Plant et al. 2001). Alternatively the intergenic transcription could mediate the formation of a silent chromatin in the absence of erythrocyte-specific transcription factors (Haussecker and Proudfoot 2005).

More recently, a lot of different techniques and approaches have been developed to study on a large scale the gene expression in different organisms. The conclusion was that the genome is widely transcribed producing many non coding transcripts of unknown function (TUFs). Most of these TUFs map in unannotated regions of the genomes previously erroneously called "junk DNA" (Mattick 2003; Cawley et al. 2004; Gustincich et al. 2006; Willingham and Gingeras 2006; Taft et al. 2007; ENCODE Project Consortium 2007).

Comparative analyses of the mouse genome (chromosomes 10,16, 17) with human chromosome 21 revealed that there is almost two times more evolutionary sequence conservation observed than expected and that these conserved sequence regions are located distal from the well-annotated exons (Dermitzakis et al. 2002). Approximately 37% (837) of the conserved sequence regions (>100 bp and >70% identity) on mouse chromosomes 10, 16, and 17 were determined to be transcribed. A series of studies based on genomic tiling arrays involving different species has been showing that a large fraction of the genome produce unannotated transcripts with reduced coding potential and of unknown function. A first attempt to evaluate the global transcriptional output of the human chromosomes 21 and 22 has shown they were one order more active than expected based on the predicted protein coding genes, suggesting the presence of many non coding transcriptional units (Kampa et al. 2004). Similar results were obtained in a tiling array-based whole genome analysis in *Arabidopsis* where more than 50% of the observed transcription came from intergenic regions and about 30% of the annotated loci were found to associate with antisense transcription

which in some cases was tissue-specific (Stolc et al. 2005; Hanada et al. 2007). Also in *Drosophila* about 40% of the transcription has been ascribed to intergenic or intronic regions of the genome, which are subject to developmentally coordinated regulation (Oliver B. 2006).

In a study aimed to evaluate the functional role of pseudogenes in humans, the pseudogene Makorin p1 was shown to be responsible for stabilizing the mRNA of Makorin1 covering a *cis*-regulatory decay-mediating sequence within the 5' region that shares homology between the two loci (Hirotsune et al. 2003). In the same study, at least 20% of the 20,000 human annotated pseudogenes was shown to be expressed, raising the question of how common may be a pseudogene-mediated regulation of gene activity (Mighell et al. 2000).

By combining chromatin immunoprecipitation (ChIP) to genomic tiling arrays, an unbiased high-resolution mapping of transcription factors binding sites (TFBS) across the genome was obtained. The binding sites of the NF-kB family member RelA/p65 were mapped across human chromosome 22, revealing that about 28% of the binding sites lie more than 50kbp away from known protein-coding genes (Martone et al. 2003). In another study, Sp1, c-Myc and p53 transcription factors binding sites along human chromosomes 21 and 22 were mapped, showing that 36% of the binding sites were within or 3' flanking to known genes and correlating with non coding RNAs, 17% of the TFBS were associated to pseudogenes or ambiguous sequences whereas only 22% of these TFBS regions were located at the 5' termini of protein-coding genes (Cawley et al. 2004).

To obtain a general picture of the active promoters along the entire human genome, a ChIP-on-chip strategy was carried out to immunoprecipitate DNA bound to TAFII250, a component of the RNA polymerase II preinitiation complex (PIC). DNA fragments were then hybridized on a tiling array containing 14.5 milion 50-mer oligonucleotides, designed to represent all the non-repeat DNA troughout the human genome. More than 10,000 binding sites were found to be representative of active promoters and 13% of these were mapping in unannotated transcription units (Kim et al. 2005).

In this case the low representation of the non coding genes among the observed binding sites could be explained by the highly conservative thresholds used to identify the TFBS and by the low occupancy by RNA polymerase of low abundance transcripts.

In a more recent analyses conducted by the RIKEN Genome Research Group for the Functional Annotation of Mouse Genome FANTOM 3 Project, it has revealed that in the mouse 34,030 full-length cDNAs clusters lack any protein coding potential and are annotated as "non coding RNA". So far, the number of non coding transcriptional units (34,030) exceeds the number of protein-coding (32,129) TUs in mouse (Carninci et al. 2005). About 63% of these ncRNAs are spliced and their expression has been proved and confirmed that they are not byproducts of genomic contamination during cloning procedures (Carninci et al. 2005, Ravasi et al. 2006). Furthermore, the vast majority of the non coding transcripts shows a very low cross-species conservation at the level of the transcribed sequence (Pang et al. 2006). Interestingly, their promoters seem having a much higher degree of conservation (5kb versus 500bp from the starting site) (Carninci et al. 2005). Furthermore, ncRNAs are transcribed starting from the 3'UTR of protein-encoding genes (Babak et al. 2005; Pang et al. 2006) suggesting some functional link between the regulation of these two classes of genes.

Widespread unannotated transcription has been confirmed by many other experimental means, including mapping 5` ends of transcripts using cap analysis of gene expression (CAGE), 3`ends by serial analysis of gene expression (SAGE), massively parallel signature sequencing (MPSS) and high-throughput full-length cDNA cloning and sequencing (reviewed in Mattick and Makunin 2006). Sequencing of 3 milion human CAGE tags from various tissues confirmed that human cells have a similar level of transcriptional diversity as revealed by the FANTOM 3 full-length cDNA collection (Carninci et al. 2005; Carninci et al. 2006; Gustincich et al. 2006).

Analysis of human gene expression by MPSS showed that more than 65% of signature sequences do not overlap with annotated transcripts, around 38% of them map to introns, 21% are antisense to known exons and 5% map to intergenic regions.

Expression analysis by SAGE tags found at least 15,000 previously uncharacterized 3'ends used in the genome, suggesting that new isoforms and genes are in the order of thousands, representing a significant portion of the genome.

A fundamental question arise from all these genome-wide observations: are all of those transcripts biologically functional or do they represent just transcriptional noise of the complex cellular biological processes? Once the new transcripts are confirmed to exist in the cell, it takes time to uncover their function. In the past 35 years seven major functional classes of ncRNAs were recognized: ribosomal (rRNA), transfer (tRNA), small nuclear (snRNA), small nucleolar (snoRNA), micro (miRNA), and Piwi-interacting (piRNA). With the exception of perhaps rRNAs and tRNAs, which were discovered in 1958, all the other classes are believed to be incomplete, and it's very likely to suppose that many other different functional classes are still to be discovered.

Many functions for ncRNAs have been identified so far, including transcriptional activation, gene silencing, chromatin epigenetic regulation, imprinting, dosage compensation, translational inhibition, modulation of protein function, binding as riboswitches to regulatory metabolites (Kiss 2002; Zamore and Haley 2005; Mattick and Makunin 2006).

The complexity of the Eukaryotic Transcriptome: Poly A+ versus PolyA- transcripts

It was 1974 when Christine Milcarek, Richard Price and Sheldon Penman at MIT found that about 30% of HeLa cell mRNA lacks poly(A) when labeled in the presence of different rRNA inhibitors. Their method of RNA fractionation precluded contamination of the poly(A)$^-$ mRNA with large amounts of poly(A)$^+$ sequences. The poly(A)$^-$ species was showed to be associated with polyribosomes, having an average sedimentation value equal to or greater than poly(A)$^+$ mRNA, and behaving like the poly(A)$^+$ mRNA in its sensitivity to EDTA and puromycin release from polyribosomes. Very little, if any, hybridization at Rot values characteristic of abundant RNA sequences between the poly(A)$^-$ RNA fractions and ^{3}H-cDNA made from poly(A)$^+$ RNA was

detected. This indicates that poly(A)⁻ mRNA does not arise from poly(A)⁺ mRNA by nonadenylation, deadenylation, or degradation of random abundant mRNA sequences. The poly(A)⁻ mRNA was stable in a long (20 hr) time period. These data indicated that poly(A)⁻ mRNA is not short-lived nuclear or cytoplasmic heterogeneous RNA contamination, and that the half-life of the poly(A)⁻ mRNA may parallel that of the poly(A)⁺ mRNA. (Milcarek, Price and Penman 1974). In a later report (Milcarek 1979) the same group showed that if poly(A)-free mRNA larger than 12 S labelled for 2 h *in vivo* is hybridized with total cellular DNA, it hybridizes primarily with single-copy DNA. When a large excess of steady poly(A)-containing RNA is added before hybridization of labelled poly(A)-free RNA, no inhibition of hybridization occurs. This indicates the existence of a class of poly(A)-free mRNA with no poly(A)+ counterpart. Some mRNA species can exist solely as poly(A)+ mRNAs. These mRNAs in HeLa cells are found almost exclusively in the mRNA species present a few times per cell (scarce sequences). Some mRNA species can exist in two forms, poly(A) containing and lacking, as evidenced by the *in vitro* translation data (Kaufmann et al. 1977). In addition, if cDNA of total poly(A)+ mRNA is fractionated into abundant and scarce classes, 47% of the scarce class cDNA can be readily hybridized with poly(A)-free mRNA. 10% of the abundant cDNA to poly(A)+ mRNA will hybridize with poly(A)-free sequences very rapidly while the other 90% hybridize 160 times more slowly, indicating two very different frequency distributions. Therefore, from these early studies the existence of three distinct mRNA classes emerged: predominantly poly(A)-free, predominantly poly(A)+ and bimorphic (Milcarek 1979).

In the following years, other studies confirmed and extended these observations demonstrating that the poly(A)-free transcripts account for a large part of the complexity of the entire transcriptome in many different systems, like, among others the mouse brain (Van Ness et al. 1979), *Drosophila* (Zimmerman et al. 1980) and tobacco leaves (Goldberg et al. 1978).

After almost 30 years, our understanding of poly(A)- transcripts has not significantly increased.

Replication-dependent histone genes were classically considered to be the only well-characterized

class of transcripts synthesized exclusively as poly(A)- transcripts (Birnstiel, Bussslinger and Strub 1985). More recently, other classes of non coding genes have been proven to be non polyadenylated (Steinmetz et al. 2001; Morlando et al. 2004; Kim et al. 2006). Several genes coding for box C/D and H/ACA snoRNAs and for the U5 and U2 snRNAs contain sequences in their 3` ends which direct the cleavage of primary transcripts without being polyadenylated (Morlando et al. 2002).

Very recently, the use of genomic tiling arrays for 10 human chromosomes allowed a global analysis of the gene expression profile at 5 nucleotide resolution, using both polyadenylated and non polyadenylated RNA fractions extracted from cytoplasm or nucleus of HepG2 cells (Cheng et al. 2005). In this study, the full-length structures of many TUFs were determined using a rapid amplification of cDNA ends (RACE) technique and resolving the RACE products on high-density arrays. Overall, they showed that there are 2.2 times as many uniquely poly(A)- (43.7%) transcribed sequences as uniquely poly(A)+ (19.4%), while the rest (36.9%) of the transcripts are bimorphic. The subcellular distribution of the transcripts revealed that a large proportion of the sequences found in the nuclear or the cytoplasmic compartments appear to be exclusive of these compartments. The amount of poly(A)+ RNA (9.7%) exclusively detected in the nucleus is less than one third of the amount of the poly(A)- transcripts (31%). Of the poly(A)+ nuclear sequences, 25% are associated to known exons, 34% map into introns, and the remaining 41% map in intergenic regions. Similarly for the poly(A)- nuclear specific sequences, 18% associate with exons, 57% with introns and 25% with intergenic tracts. Importantly, 10.6% of the transcripts exclusively present in the nucleus are bimorphic. In summary, 75% of the polyadenylated nuclear transcripts and 82% of the poly(A)-free nuclear RNAs are currently unannotated.

Poly(A)+ (3.1%) sequences exclusively detected in the cytosol are less than half as abundant as poly(A)- (6.5%) cytosolic sequences. Bimorphic transcripts esclusively detected in the cytosol are the 0.6% of the total. Of the poly(A)+ transcripts exclusively in the cytosol, 43% are associated with known exons, 22% with introns and the remaining 34% with intergenic regions. Whereas regarding the poly(A)- cytosolic RNA fraction, the distribution is 16% mapping to exons, 36% to

introns and 48% allocated into the intergenic tracts. In total, 56% of the polyadenylated cytosolic transcripts and 84% of the poly(A)-free cytosolic transcripts result to be without any annotation. Taken together, these observations show that almost 80% of the non polyadenylated transcripts and a percentage between 56 and 75% of the polyadenylated transcripts are not yet characterized, with the introns and the intergenic non coding regions being the most rich source of unknown genes (Cheng et al. 2005).

The complexity of the Eukaryotic Transcriptome: Nuclear versus cytoplasmic transcripts

Till only few years ago, it was generally accepted that in a typical mammalian cell, about 14% of the total RNA is present in the nucleus (Alberts et al. 1994) and about 80% of this nuclear fraction of RNA was believed to be processed before leaving for the cytoplasm. The other 20% was estimated to be composed of snRNAs and snoRNAs, playing an active role in the processing events. Some of these molecules had been in the cytoplasm where they were coated with different proteins before being transported back into the nucleus. Recent experimental studies dramatically changed this view, reporting that, depending from the cell line, between 40% and 50% of all the transcripts remain into the nucleus as can be detected only in this compartment, whereas between 10% and 15% are exclusively retrieved in the cytosolic fraction. The remaining part (35-50%) are shared between the nucleus and the cytoplasm (Cheng at al. 2005; Kapranov et al. 2007). By the use of tiling arrays covering all the human genome at 5bp resolution, Kapranov et al. identified three major classes of RNAs: promoter-associated shortRNAs (PASRs), termini-associated shortRNAs (TASRs) and promoter-associated longRNAs (PALRs). Interestingly, a significant fraction of shortRNAs (44%) overlapped longRNA transfrags and showed a significant enrichment for evolutionary conserved sequences. Most of the longRNA transfrags overlapped to 5' boundaries of protein-coding genes but only around the first exon-intron region. 39% of the PASRs and 35% of

the TASRs showed a syntenic conservation between mouse and human. Overall, these RNA maps provided a virtual genealogy of the transcripts in the cells, suggesting that an appreciable fraction of protein-coding genes are expressed only in the first exon and intron, and that transcription may have different states characterized by different lenghts of the transcripts. The ultimate fate of the RNAs derived from a particular locus could be predicted based on their retention in the nucleus, transport to cytosol or processing into shortRNAs. The functional role of these shortRNAs covering both 5'start sites and 3'termination sites could be related to chromatin modifications. This is also suggested by other studies in which unstable longRNAs were postulated to be involved in regulation of gene expression (Davis et al. 2006; Martianov et al. 2007). Most of mammalian RNA molecules undergo extensive processing and the "left-over" RNA fragments (excised introns and RNA sequences 3′ to the cleavage/poly-A site) are completely or partially degraded in the nucleus. Incompletely processed and otherwise damaged RNAs are also eventually degraded in the nucleus as part of the quality control system of RNA production. This degradation is carried out by the exosome, a large protein complex that contains, as subunits, several different RNA exonucleases. The export of RNA molecules from the nucleus is delayed until processing has been completed. Therefore, any mechanism that prevents the completion of RNA splicing on a particular RNA molecule could in principle block the exit of that RNA from the nucleus. This feature forms the basis for one of the best understood examples of "regulated" nuclear transport of mRNA, which occurs in HIV. The virus encodes a protein, called Rev, that binds to a specific RNA sequence (called the Rev responsive element, RRE) located within a viral intron. The Rev protein interacts with a nuclear export receptor (exportin 1), which directs the movement of viral RNAs through nuclear pores into the cytosol despite the presence of intron sequences. The only way for RNAs to leave or enter the nucleus is via one of the many nuclear pore complexes that cover the nuclear membrane. Nuclear pore complexes are highly organized structures that play an active role in movement of molecules into and out of the nucleus (Wente 2000). As in many biochemical systems, the energy is obtained by hydrolysis of one of the high-energy phosphate-phosphate bonds

in a ribonucleotide triphosphate, in this case by converting GTP to GDP. Energy generation is carried out by a protein called Ran, and transport requires receptor proteins called "karyopherins", or exportins and importins depending on the direction of their transport activity. There are at least 20 different human karyopherins, each responsible for the transport of a different class of molecule - mRNA, rRNA, etc. Examples include exportin-t that has been identified as the karyopherin for export of tRNAs in yeasts and mammals. Transfer RNAs are directly recognized by exportin-t, but other types of RNA are probably exported by protein-specific karyopherins which recognize the proteins bound to the RNA, rather than the RNA itself. This also appears to be the case for import of snRNA from cytoplasm to nucleus, which makes use of importin β, a component of one of the protein transport pathways (Nigg 1997; Weis 1998). Export of mRNAs is triggered by completion of the splicing pathway, possibly through the action of the protein called Yra1p in yeast and Aly in mammals (Zhou et al. 2000; Keys and Green 2001). Once outside the nucleus, there are mechanisms that ensure that mRNAs are transported to their appropriate places in the cell. It is not known to what extent protein localization within the cell is due to translation of an mRNA at a specific position or to movement of the protein after it has been synthesized, but it is clear that at least some mRNAs are translated at defined places. For example, those mRNAs coding for proteins that are to be transferred into a mitochondrion are translated by ribosomes located on the surface of the organelle. It is assumed that protein 'address tags' are attached to mRNAs in order to direct them to their correct locations after they are transported out of the nucleus. However, very little is known about this process. When mRNAs encode a protein that is destined to be secreted or expressed on the cell surface, it will be directed to the endoplasmic reticulum (ER) by a signal sequence at the protein's amino terminal. Components of the cell's protein-sorting apparatus recognize the signal sequence as soon as it emerges from the ribosome and direct the entire complex of ribosome, mRNA, and nascent protein to the membrane of the ER, where the remainder of the polypeptide chain is synthesized. In other cases the entire protein is synthesized by free ribosomes in the cytosol, and signals in the completed polypeptide chain may then direct the protein

to other sites in the cell. Some mRNAs are themselves directed to specific intracellular locations before translation begins. Presumably it is advantageous for the cell to position its mRNAs close to the sites where the protein is required. The signals that direct mRNA localization are typically located in the 3′ untranslated region (UTR) of the mRNA molecule. A striking example of mRNA localization is seen in the *Drosophila* egg, where the mRNA encoding the bicoid gene regulatory protein is localized at the anterior tip of the developing egg by attachment to the cytoskeleton. When the translation of this mRNA is triggered by fertilization, a gradient of the bicoid protein is generated that plays a crucial part in directing the development of the anterior part of the embryo. Many mRNAs in somatic cells are localized in a similar way. The mRNA that encodes actin, for example, is localized to the actin-filament-rich cell cortex in mammalian fibroblasts by means of a 3′ UTR signal. The 5′ cap and the 3′ poly-A tail are necessary for efficient translation of protein-coding genes, and their presence on the same mRNA molecule thereby signals to the translation machinery that the mRNA molecule is intact. The 3′ UTR often contains a "zip code," which directs mRNAs to different places in the cell. mRNAs also carry information specifying the average length of time each mRNA persists in the cytosol and the efficiency with which each mRNA is translated into protein. In a broad sense, the untranslated regions of eucaryotic mRNAs resemble the transcriptional control regions of genes: their nucleotide sequences contain information specifying the way the RNA is to be used, and proteins that interpret this information bind specifically to these sequences. Thus, over and above the specification of the amino acid sequences of proteins, mRNA molecules are rich with many additional types of information. Thus a cell can control gene expression by controlling when and how often a given gene is transcribed (**transcriptional control**), controlling how the RNA transcript is spliced or otherwise processed (**RNA processing control**), selecting which mature mRNAs in the cell nucleus are exported to the cytosol and determining where in the cytosol they are localized (**RNA transport and localization control**), selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**),

or selectively activating, inactivating, degrading, or compartmentalizing specific protein molecules after they have been made (**protein activity control**). For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points, only transcriptional control ensures that the cell will not synthesize superfluous intermediates.

The complexity of the Eukaryotic Transcriptome: What's the function of many repeat elements?

About 98% of the mammalian genomes is composed of DNA of unknown function and for this reason is also known as "junk DNA". Most of this junk DNA is composed of repetitive or semirepetitive DNA: up to 45% of the human genome is covered by transposable elements, a result of over 3 milion SINE- LINE- DNA- LTR-transposon insertion events (Lander et al. 2001). Their average length is about 400bp, ranging from 0.1 up to 8kbp, with an average distance between them of 476bp (Simons et al. 2006). A similar transposon density is also observed in mouse, although most of transposons present in human and mouse have entered these two lineages independently since their divergence (Waterston et al. 2002). In the past few years a sort of molecular gold rush has being started to uncover the hidden functional elements present in this huge unexplored part of the genomes, which can provide some funtional clues about how the genomes operate and change through time. Since they were discovered by Barbara McClintock about 50 years ago, they were called "control elements", due to the fact they were able to influence the expression of neighbouring genes (McClintock 1956), and in 1969 Roy Britten and Eric Davidson suggested that they can provide the means for fortuitous evolutionary innovations. In their theory, new branches on the tree of life and more complex organisms arose, at least in part, from changes in the way genes were regulated. They also argued that these changes were often caused by repetitive elements, many of which were later identified as transposons. Sets of genes are turn on at particular time and specific places during development, and transposons may copy themselves into different parts of the genome and occasionally reconfigure these controls. Bringing, for instance, two independent gene networks under a common regulatory control, changing position along the genome, transposons can

contribute to generate new cell types and new structures (Britten and Davidson 1969). Over the years this idea lost momentum, also due to many difficulties to be tested. But after 35 years, when researchers started to compare genomes, it became clear that not all the junk DNA was junk and also the idea of Britten and Davidson started to be revaluated. The discovery that 481 stretches of at least 200bp were exactly the same in human, rat, mouse, chicken, dog, and to a less extent fishes, led to define a set of "ultraconserved elements" (Bejerano et al. 2004). Three-quarters of these sequences lie outside protein-coding genes, particularly nearby genes involved in the regulation of transcription and development.

The remaining ones are most often located either overlapping exons in genes involved in RNA processing or in introns. And even more recently, Venkatesh and collaborators analyzed the draft genome of the elephant shark comparing with human genome, and found that these two species separated 350 milion years ago share 4800 conserved sequences. The conserved sequences tend to cluster near genes coding for DNA binding and transcription regulatory proteins (Venkatesh et al. 2007). According to other studies, many conserved non coding elements appeared for the first time with the evolutionary division between lampreys and sharks, suggesting that these sequences could have been playing a key role in the establishment of the vertebrate development gene-regulatory networks (Woolfe and Elgar 2007).

A deeper analysis of these conserved sequences has revealed that some of them are transposable elements. Okada Norihiro and collaborators have sequenced the coelacanth Lathimeria, a "living fossil" species which dates back 400 milion years ago, and found a couple of short interspersed repetitive elements (SINE) which are enough conserved among them to have a common origin. Then they searched other genomes for the presence of similar sequences and found that in sea urchin, zebrafish, catfish, trout, salmon, hagfish, dogfish shark and amphioxus a common set of SINEs were conserved. They called this superfamily of repetitive sequences DeuSINEs, as they were commonly present in all Deuterostomia. This group is actually composed of five families of tRNA-derived SINEs, in which a 5S rRNA-derived promoter is flanked by a highly conserved core

(Deu domain) and a 3`tail (Nishihara, Smit and Okada 2006). The same authors also found about 1000 copies of a subset of these SINEs called AmnSINE1 (Amniota SINE1) which are conserved from humans to birds, and some copies date back before the mammalian-bird split about 310 milion years ago. Also AmnSINE1 share the common chimeric structure of a 5S rRNA-derived promoter and a tRNA-derived central core, previously seen in the DeuSINE superfamily. Out of 1000 copies in the human genome, 105 map to loci evolutionary highly conserved among mammalian orthologs, being good examples of a transposable element of which a significant fraction of the copies have acquired and maintained genomic functionality (Nishihara, Smit and Okada 2006). In an independent study, Xie and collaborators reached similar conclusions. In a search for DNA sequences present in the human genome in multiple copies, they found one element which showed a high homology with the 180 bp core of the active transposon in Zebrafish SINE3 family, and is present in 124 different genomic locations. Surprisingly this core sequence is retrieved in the same places in many other genomes, including few copies in the coelacanth. Their family of conserved SINEs has been proved to be the same discovered by the group of Okada (Xie, Kamal and Lander 2006). They also identify additional 95 families of conserved noncoding elements (CNE) which likely predate the mammalian radiation, bringing attention to the importance of both transposons and other types of non coding repetitive elements in the functional organization of the genome. Bejerano and collaborators found that among the ultraconserved elements is present also a family of transposons which was first discovered in 10,000 copies in the coelacanth and called "LF-SINEs" were LF stands for "lobe –finned" fish. The same family of SINEs was also found in all the tetrapods genomic sequences available in databases (Bejerano et al. 2006). They demonstrated that some of these transposons are likely to function as enhancers to control the spatial and temporal expression profile of specific set of genes. In particular, out of 167 conserved elements tested in a transgenic mouse enhancer assay, 45% showed to have a reproducible tissue-specific enhancer activity at embryonic day 11.5. Furthermore, the majority were directing the expression in various regions of the central nervous system. Recently, another study support with experimental evidence

the importance of retrotransposition events in neuronal differentiation. Muotri and colleagues studied the L1 retrotransposition events in a transgenic mice concluding that L1 retrotransposons can alter the expression of neuronal genes and produce somatic mosaicism in the neuronal precursor cells (Muotri et al. 2005).

Interestingly, a survey of the Opossum genome searching for conserved non coding elements (CNE) led to identify some CNEs specific of placental mammals and diverged from marsupials. Up to 16% of these eutherian- specific CNEs were mapped to at least one of 12 different families of transposable elements. On contrary, the eutherian CNEs that are also present in Opossum only rarely overlap to recognizable transposable elements (0.7%) (Mikkelsen et al. 2007). Those CNEs are likely to be derived from transposable elements, as is difficult to recognize the future of a transposon copy in the genome that appeared above 100-200 milion years ago. The authors of this study concluded that transposable elements were likely instrumental for regulatory changes underlying features characteristic of placental mammals. Overall it seems that conserved repetitive DNA in the genome, at least in part overlapping with known transposons, have retained a function which has not been eliminated by the strong negative pressure to which non coding conserved elements are usually subjected (Ashtana et al. 2007; Kamal et al. 2006; Ponjavic et al. 2007)

The complexity of the Eukaryotic Transcriptome: S/AS pairs

Detection of potential sense-antisense (S/AS) pairs has contributed to increase our description of the transcriptome. While previous analyses of the mammalian transcriptome suggested that about 20% of protein encoding genes have sense-antisense transcription (Okazaki et al. 2002; Kiyosawa et al. 2003; Yelin et al. 2003; Chen et al. 2004; Werner and Berdal 2005), SAGE indicated that at least 50% of all the transcripts have a corresponding antisense transcript, even considering the brain alone (Siddiqui et al. 2005). CAGE tags combined with a detailed analysis of mouse full-lenght

cDNAs shows that up to 72% of the transcriptional units exhibit S/AS transcript pairs (Katayama et al. FANTOM Consortium and RIKEN Genome Exploration Research Group 2005).

Out of 43,553 total genome-mapped non redundant transcriptional units, 72% overlap with some cDNA, 5' or 3' EST sequence, or CAGE tag or PET tag on the opposite strand. In mouse, S/AS pairs are found in 87% of the protein coding TUs (20,714) and in 58,7% of the non coding TUs (22,839). Even considering only S/AS pairs supported by full-length cDNA overlap, for 4,520 TUs S/AS pairs were detected on the exons while 4,129 TUs have S/AS pairs overlapping to non exonic regions, suggesting that immature RNAs in the nucleus and introns can originate smaller RNAs with biological activity (Mattick and Makunin 2005).

Furthermore, the antisense ncRNAs expression level could be still underestimated, since random primed CAGE libraries showed to have a larger content of antisense ncRNAs compared to oligodT-primed CAGE libraries (Katayama et al. 2005).

This is concordant with the observation that antisense transcripts are poorly polyadenylated (Kiyosawa et al. 2005) and suggest that S/AS pairs could involve poly(A)- RNAs as well as long non coding transcripts. Antisense transcription is very often associated to imprinted loci, as more than 81% of the imprinted TUs show S/AS pairs, when sequences antisense to introns are taken into account. S/AS pairs are unevenly distributed along the genome, as mouse chromosomes 4 and 17 clearly show a S/AS pairs density larger than average while chromosomes 6, 9 and 13 less frequently contained them. The X chromosome has the fewest S/AS pairs, probably related to the monoallelic inactivation, due to the strong association of the antisense transcription to the imprinted loci. Interestingly, the mouse underrepresented chromosome 6 is mostly homologous to the human chromosome 7, which is known to be rich in genes transcribed by RNA polymerases I and III, which may be not captured by the approach used in this study (Katayama et al. 2005).

A differential expression of S/AS pairs is noticed among different tissues and conditions. A preference for certain types of genes is also showed as summarized by GO terms distribution, with a clear overrepresentation of cytoplasmic proteins and a significant underrepresentation of genes

encoding for membrane and extracellular proteins. This suggests that S/AS pairs are subjected to a specific regulation and unlikely represent transcriptional noise (Katayama et al. 2005).

Overlap of *cis*-S/AS pairs can involve different portions of the corresponding TUs, giving rise to three basic types of configurations: "head-to-head" or divergent in which the TUs overlap at their 5'ends, "tail-to-tail" or convergent in which the TUs overlap at their 3'ends, and fully overlapping. Although various studies (Lehner et al. 2002; Yelin et al. 2003; Chen et al. 2004, 2005) suggest that the most common S/AS transcripts overlap as tail-to-tail, CAGE tags suggest that head-to-head overlap is more frequent than convergent antisense (Katayama et al. 2005).

Monitoring their expression in a time-course experiment of macrophage cells activation by lypopolysaccharide, most of the S/AS pairs showed a proportional co-regulation in response to the stimulus, whereas only a minority showed a reciprocal regulation, with increased level of the sense transcript corresponding to a decreased level of the antisense transcript.

Although concordant regulation is observed more frequently, there are few well characterized examples in which the two transcripts are expressed reciprocally. For instance, a targeted siRNA-mediated inhibition of the transcript of Ddx39 led to an increase in the level of CD97 mRNA, but the reciprocal effect was not observed (Katayama et al. 2005).

When a direct correlation was observed, a decrease of sense hypothetical aminoacyl-tRNA synthase class II-containing protein resulted in decreased antisense C/EBP delta expression. The reverse was not found. A positive correlation was also noticed when a non coding RNA was the partner in the S/AS pair.

All these evidences argue against a simplistic assumption of a negative regulatory role of antisense transcription. S/AS pairs can potentially provide a dsRNA intermediate, which is the substrate of the enzyme Dicer, the main player of RNA interference (Ambros 2004). Dicer is involved in the production of small interfering RNAs (siRNAs), which direct the cleavage of the complementary mRNA. siRNAs can also take part in the transcriptional gene silencing in the nucleus (Matzke and Birchler 2005), while Dicer has been involved in the heterochromatin formation in vertebrates

(Fukagawa et al. 2004). When non coding antisense RNAs overlap the sense promoter, they seem to play a role in the local chromatin modification or methylation, (Imamura et al. 2004; Murrell et al. 2004; Andersen and Panning 2003).

To summarize, in 30% of the cases in which experimental validation after perturbation was carried out, S/AS pairs showed both a negative and a positive regulation of each other. The existence of both positive and negative correlations in the levels of S/AS pairs transcription suggest that mechanisms other than siRNA-like mode of action account for the antisense transcripts function in the genome.

The complexity of the Eukaryotic Transcriptome: the impact of the technical limitations on gene discovery and genome annotation

Recently, many technological advancements have increased our ability to study the mammalian genome and transcriptome. Among them, the shotgun sequencing strategy of expressed sequence tags (EST) together with *in silico* computer-assisted assembly of sequences, many different genome projects have been very important.

The development of full-length cDNA libraries (reviewed in Das et al. 2001), together with the appearance of cDNA microarrays (Schena et al. 1995), oligonucleotide gene chips (Kane et al. 2000) and tag-based platforms (Harbers and Carninci 2005), made possible to start a fine characterization of the output of the trascriptome. In parallel, the development of other relatively high-throughput methods as yeast two- and three-hybrids system (Uetz et al. 2000), phage display libraries (Cwirla et al. 1990; Scott and Smith 1990), chromatin immunoprecipitation (ChIP) (Orlando, Strutt and Paro 1997), ChIP on chip (Ren et al. 2000), 2D electrophoretic separation coupled with mass-spectrometry (Mann and Wilm 1995) have greatly enhanced the analysis of the proteome output, giving a more global view of the network of protein-protein and protein-DNA interactions within the cells. All these different approaches are producing huge amount of data,

usually stored in relational databases, which demanded for new bioinformatic algorithm and infrastructures to extract useful informations and integrate different data sets into a functional description of cellular activities.

Despite many progresses, it has been evident that we are still further away from having a comprehensive and exhaustive view of how molecular biological systems work, and researchers worldwide just started to uncover the tip of an iceberg, as many limitations still prevent them from having a deep integrated view of the dynamics that are generated and maintained inside the cells.

One of the most relevant limitation is represented by the minimum scale to which biological investigations of gene expression can reliably be conducted. Suddenly the importance of cellular diversity came out as one of the most striking features accounting for the biological complexity of higher eukaryotes. During the past decades, many studies have been focused on single cell types, often describing single aspects of the biology of those cells, but with the advent of the post-genomic era, there is a growing need for developing reproducible methods to analyze gene expression of single cells in an unbiased high-throughput way.

These small scale-technologies are needed since the expression profiles obtained from complex tissues are likely to miss most of the complexity of the transcriptome, which is largely composed of low expressed cell-specific transcripts.

Furthermore, the emerging picture furnished by the current methodologies is more or less a static snapshot of the gene activities in the cell, still far away from being representative of what is the true dynamical status of a cell.

To achieve this goal, a great contribution it would probably come from the development of new "real-time" techniques, capable of detecting and monitoring the global variations in gene expression in real time, taking into account all the genes at the same time.

Importantly, new approaches must aim to the analysis of expression in sub-cellular compartments or different nuclear bodies.

Several advancements have already opened the way in this direction: live molecular imaging studies (reviewed in Ottobrini et al. 2006) made possible for instance to monitor single neurons activation in real time, while chromosome conformation capture technology (3C) rendered available a tool for analyzing the chromatin arrangement into the nucleus over long distances (Dekker et al. 2002), giving a preliminary view of the higher-order chromatin domains.

The future development of a 4C technique, which can combine the chromosome capture conformation assay to DNA microarrays or tag-based technologies, would further help to uncover the links between nuclear architecture and genome activity.

More traditionally, cell-biologist`s approaches like multicolour DNA and RNA fluorescence in situ hybridization (FISH) (Carter 1994) or three – four-dimensional light microscopy have been also employed to complement the molecular approaches, although these techniques are currently limited by the Abbey limit for the maximum resolution.

Even focusing only on gene discovery and genome annotation, a series of limitations prevented access to large part of the transcriptome, as many transcripts are expressed in a cell-specific manner and at low level, and many of them are thought to be poorly polyadenylated.

Traditional cDNA libraries are made by priming the transcripts with oligodT and yield on average a 20-30% of full-length molecules (Marra et al. 1999). Starting from the '90s several methods have been developed to enrich for full-length cDNAs up to 90%, in order to clarify the number and the structure of protein-coding genes, including the untranslated regions (Das et al. 2001).

These libraries became a standard for high-quality large-scale sequencing projects due to the fact that they yield the full-length sequence data at a fraction of the cost of the entire genomic DNA sequencing. The full-length library construction took advantage of the presence of a cap structure at 5`end of the genes transcribed by RNA polymerase II. This is usually represented by an inverted 3`-5` pppGm[7] nucleotide, which is added at the 5`end at very early stage of RNA synthesis, and may serve to control mRNA localization, mRNA stability and translation efficiency according with the mRNA factory model (Miura 1981; Bentley 2002).

At least three main systems were employed to synthesize full-length cDNA libraries: (1) SMART^R which makes use of strand-switching activity of a MMLV reverse transcriptase to add a GGG containing oligonucleotide at the 5`end and selectively recover the full-length molecules, subsequently amplified by PCR (Zhu et al. 2001); (2) oligo-capping method, in which uncapped RNAs are dephosphorylated by a phosphatase whereas all the capped mRNAs are treated with tobacco acyd pyrophosphatase to remove the cap and leave a 5` phosphate group, to which an oligonucleotide is covalently added by RNA ligase, followed by library preparation and PCR amplification (Maruyama and Sugano 1994) ; and (3) the cap-trapper method, which after oxidation of the diol group of the cap site with NaIO4 and a biotinilation with a long-arm biotin hydrazide, makes use of RNAse I to cleave and remove truncated and incomplete cDNA/mRNA hybrids, with a good efficiency, without need for an additional PCR.

Importantly, all of these methods have to optimize synthesis of full length cDNA for long transcripts.

To overcome this problem, lambda vectors specifically designed for long cDNA cloning were employed to host up to 37.5 kb long insert, yielding libraries with up to twofold larger sequence diversity compared to libraries of shorter size (Carninci et al. 2001).

Furthermore, to increase the gene discovery rate by sequencing randomly selected cDNA clones, removal of undesired common and highly expressed sequences was achieved by normalization and subtractive hybridization (Bonaldo et al. 1996). On average, few genes account for 20-30% of the total mass of mRNAs, while intermediately expressed genes (about 1000-2000 distinct transcripts) and rarely expressed species (which constitute the large majority of the transcripts) constitute about 30-50% and 30-40% respectively of the total cellular mRNAs.

This implies that in order to have a better coverage and avoid prohibitive sequencing costs it is mandatory to enrich the libraries for rare transcripts using subtractive hybridization strategies.

Subtraction and normalization have been widely used in many EST projects to enrich for novel transcripts in various organisms (Hillier et al. 1996; Marra et al. 1999; Sheetz et al. 2004) as well as for full-length cDNA-based EST libraries (Carninci et al. 2003).

Despite the great increase in the discovery rate, the subtraction and normalization negatively select against alternative splicing forms, which makes those libraries unsuitable for studying alternative splicing isoforms. On the other hand it has been noticed that full-length cDNA libraries still represent one of the best tool to characterize alternative splicing events, due to the fact that other methods do not usually yield the entire structure of the transcripts but interrogate only parts of it, like in the case of genomic tiling arrays.

To circumvent this limitation, new approaches in the library construction are needed, which could take advantage from the selection of mis-paired nucleic acid hybrids as suggested from recent studies (Watahiki et al. 2004; Thill et al. 2006).

Besides displaying single alternative exons, new methods should include full-length structure determination, because is not possible to reconstruct the structure of full-length mRNA without full-length cDNAs (Carninci 2007).

Another limitation in the subtraction and normalization of cDNA libraries derives from the use of dsDNA drivers for subtraction which may result in under-representation of S/AS transcripts in the library as demonstrated by human and rat EST non full-length libraries (Bonaldo et al. 1996).

Technologies to unveil transcriptome complexity: cDNA microarrays, oligonucleotide Gene Chip and tiling arrays

Array technologies take advantage of the universal inherent property of all the complementary nucleic acid molecules to hybridize to each other in a highly specific way.

The traditional solid-phase array is a collection of microscopic DNA spots attached to a solid surface, such as glass, plastic or silicon chip. The affixed DNA segments are known as probes, thousands of which can be placed in known locations on a single DNA microarray. Microarray

technology evolved from Southern blot, whereby fragmented DNA is attached to a nitrocellulose membrane and then probed with a known labeled gene or fragment. In spotted microarrays (or two-channel or two-colour microarrays), the probes are cDNA, oligonucleotides or small fragments of PCR products that correspond to mRNAs and are spotted onto the array surface. This type of array is typically hybridyzed with cDNA derived from two samples to be compared, labelled with two different fluorophores.

Fluorescent dyes commonly used for cDNA labelling include Cy3, (fluorescence emission at 570 nm), and Cy5 (fluorescence emission at 670 nm). The two Cyanine-labelled cDNA targets are mixed and hybridized to a single microarray, then scanned to visualize fluorescence of the two fluorophores after excitation with a laser beam of a defined wavelength. Relative intensities of each fluorophore may then be used in a ratio-based analysis to identify up-regulated and down-regulated genes. Absolute levels of gene expression cannot be determined in two-channel microarrays, but relative differences in expression among different spots can be estimated with some oligonucleotide arrays.

Since the first report of application in 1995 to a study conducted in *Arabidopsis* (Schena et al. 1995), cDNA microarrays have been progressively established as a high-throughput platform for gene expression profiling, due to the growing availability of cDNA collections from various organisms.

Several improvements and technical advancements have been carried out, from the screening of large tissue preparations to the potential analysis of small number of cells, with the support of a variegated set of amplification techniques, which operate either at the level of the signal detected or in the sample preparation.

A step forward in the customization of the microarrays as platforms for gene expression profiling has been represented by the advent of oligonucleotide based probes, because these do not rely from already existing cDNA collections and can be designed based on *in silico* annotated genes present in publicly available genomic sequences.

In a pioneer report (Kane et al. 2000), the validity of 50-mer oligonucleotide probes spotted on a glass slide was tested in comparison to adjacent PCR products having the complementarity for the same targets. No significant difference in sensitivity between oligonucleotide probes and PCR probes was observed and both had a minimum reproducible detection limit of approximately 10 mRNA copies/cell. Specificity studies showed that for a given oligonucleotide probe any 'non-target' transcripts (cDNAs) >75% similar over the 50 base target may show cross-hybridization. Thus the authors concluded that non-target sequences which have >75-80% sequence similarity with target sequences will contribute to the overall signal intensity.

Further improvements in the array technologies became possible with the advent of more sophisticated synthesis and assembly techniques. Originally developed by Affymetrix in 1994 (Pease et al. 1994; Chee et al. 1996), the photolitographic process, typically used in the electronic microcircuits production, has been applied for a light-assisted oligonucleotide synthesis, which allows a high-throughput and automated pipeline for the production of the, so called, gene chips. These microarrays give estimations of the absolute value of gene expression and therefore the comparison of two conditions requires the use of two separate microarrays. Currently, commercially available oligonucleotide-arrays can be either produced by piezoelectric deposition with full length oligonucleotides or *in-situ* synthesis. Long oligonucleotide-arrays are composed of 60-mers, or 50-mers and are produced by ink-jet printing on a silica substrate. Short oligonucleotide-arrays are composed of 25-mer or 30-mer and are produced by photolithographic synthesis (Affymetrix) on a silica substrate or piezoelectric deposition (GE Healthcare) on an acrylamide matrix.

One of the major drawback of the cDNA and oligonucleotide microarrays is represented by the fact that the output of expression profiles is completely hypothesis-bound, since they rely for the probe design from the currently known gene annotations.

More recently Affymetrix has developed GeneChip Tiling Arrays, which can be used to interrogate genomes at regular intervals, including both annotated and so called "junk" regions. The genomic

tiling arrays have been employed as discovery tool to identify novel transcriptional elements and evaluate antisense transcription (Kapranov et al. 2002; Rinn et al. 2003; Chen et al. 2005).

In a study in which tiling arrays with 25-nt probes at 35bp resolution have been used to interrogate human chromosomes 21 and 22, 26.5% of the probes were positive in at least one out of 11 human cell lines using poly(A)+ RNA for target preparation (Kapranov et al. 2002).

In a more wide study using 5bp resolution tiling arrays covering 10 human chromosomes, accounting for 43% of the human genome, 10.1% of the probes were positive in at least in one cell line and 4.9% were common to all of them (Cheng et al. 2005).

Another study using tiling arrays at the same resolution (25nt probes in 5bp spacing), and interrogating the same human chromosomes revealed that in the case of HepG2 liver cells, when poly(A)+ and poly(A)- RNA from the nucleus and the cytoplasm was used as target, 15% of the probes were positively detected and 41.5% of the transcripts were found to be in the nuclear fraction. The Yale group, using tiling arrays containing all the human chromosomes (36nt probes at 46bp spacing) and screening poly(A)+ pooled liver RNA, reported of 10,595 novel transcription active regions (TARs) which were not previously annotated, and 14,884 transcripts were predicted by Genscan without any other cDNA or EST support.

Even more recently, genomic tiling arrays in conjunction with other methods have been applied in the ENCODE project to characterize the funtional elements of the transcriptome for a selected portion of the genome (ENCODE project consortium 2007).

However, despite the great results in gene expression profiling achieved by challenging microarrays technologies, the drawbacks of these systems remain many. The major limitations are represented by: (1) dependence of the platform design from the experimental system subjected to investigation, (2) inherent noise related to the hybridyzation kinetics and to the fluorochromes signal amplification (3) no clear distinction between different transcript isoforms as alternatively spliced transcripts (4) no direct way to analyze the regulatory regions of the transcriptomes like alternative transcription initiation or termination sites (5) limitations in the detection of rare transcripts as 25-nt

Affymetrix GeneChips, the basis of tiling arrays, was reported to detect <55% of the rare transcribed mRNAs (Draghici et al. 2006).

Technologies to unveil transcriptome complexity: Serial Analysis of Gene Expression – SAGE

SAGE has opened the field of expression profiling to the use of short sequence tags (Velculescu et al. 1995). Its striking success depends on the ability to provide absolute expression values without need for any probe design, which proved to be instrumental for the quantitative analysis of biological samples. The concept is based on cleaving a cDNA with a so called "anchoring enzyme", commonly NlaIII, to create a recognition site for a class IIs restriction endonuclease adjacent to the cloning site. This class of restriction enzymes cleave outside of their recognition sites, and those cutting 14-20 bp away from their binding sites are particularly useful for tag preparation.

This allows for the isolation of short sequencing tags at locations defined by the distribution of the anchoring enzyme recognition sites along the genome.

Both 5'end-related and 3'end-related sequences can be isolated and cloned into concatamers by the classical SAGE methodology, but the great majority of the published studies have focused only on the tags from the 3'ends of transcripts.

Long SAGE (MmeI; 20bp tags) (Saha et al. 2002) and SuperSAGE (EcoP15I; 26bp tags) (Matsumura et al. 2003; Matsumura et al. 2008) can produce tags which are long enough to be unambiguously mapped to the genome since employing different class IIs endonucleases and other modifications of the protocol.

In the case of classical SAGE, a virtual reference tag dataset can be constructed by scanning publicly available mRNA sequences for the presence of the 3'-most anchoring restriction site, followed by the extraction of the sequence downstream of that site.

Subsequent transcript identification is performed matching the virtual tags in the reference database to the experimentally obtained tags (Lash et al. 2000).

In the case of LongSAGE and SuperSAGE experimentally obtained tags can directly be alligned to the genomic sequences for the assignment of their locations.

A statistical analysis of the tag frequencies is required to compare the expression values of different samples and to assess how the noise affects method-to-method or sample-to-sample variability (Powell 2000; Man, Wang and Wang 2000; Wang 2005). Some tags can ambiguously match to several different genomic locations, for instance due to the presence of repetitive elements in 3'UTRs, which could result particularly problematic for 3'end-related tags. Depending on the possibility to be confirmed by means of other independent experimental approaches, the mapping strategy has to be evaluated case by case.

In general for high-throughput studies, only stringent mapping conditions ensure satisfying results.

Even at risk of losing some information, mapping should always be tested to assure high quality and reliable output data. "Orphan SAGE 3'-tags" that cannot be unambiguously assigned to any genomic locations, may derive from internal priming in the cDNA synthesis, incomplete digestion with the anchoring enzyme, alternative polyadenylation, or even more in some rare cases from splicing events near the 3'end of transcripts.

Furthermore, single nucleotide polymorphisms (SNPs), which occur every 100-300bp in the human genome (Wang et al. 1998; Sachidanandam et al. 2001), may result in different sequences and/or in the disruption of the recognition sites for the anchoring enzymes.

One big limitation of the SAGE technique is its relying on the presence of the "anchoring enzyme" recognition site, since a transcript lacking this signal in the 3'end or 5'end-related sequences cannot be detected.

Another big limitation of the SAGE approach is that this method cannot be used to directly identify and isolate regulatory regions, which limits understanding of the logic behind the transcriptome.

The great advantages are represented by the fact that SAGE provides absolute levels of gene expression without need for any probe design, and the great sensitivity, which has been proved to

be 26-times more sensitive than EST methods to detect low-abundance and rare transcripts (Sun et al. 2004).

Technologies to unveil transcriptome complexity: Cap Analysis of Gene Expression – CAGE

Among the currently available gene expression profiling methods, the Cap Analysis of Gene Expression (CAGE) technique is the most relevant sequencing-based approach for extracting biologically relevant informations about transcriptional regulatory regions..

CAGE allows direct transcription start sites (TSS) identification, isolation and cloning of the 5'end-related tags, providing useful informations about promoters activity (Shiraki et al. 2003).

Since TSS cannot be accurately predicted by computational means, the determination of 5'end sequences of transcriptional units, responsible for their regulation, entirely depends on strategies to map experimentally obtained tags to genomic sequences.

The concept of CAGE is based on capturing the true 5'ends of transcripts by using the "cap-trapping" technology (Carninci et al. 1996), which takes advantage of the presence of the cap structure for binding and selection of full-length cDNA molecules. After the partially truncated cDNA, which do not reach the starting site, are discarded by using RNAse H to degrade the RNA in hybrid duplexes, an adaptor containing a recognition site for a classIIs restriction enzyme (MmeI) is ligated to the 5'end. As result of the digestion with the restriction enzyme, 18bp 5'fragments are released and, after the ligation of a second adaptor containing a different restriction site, the 20 bp long 5'end-tags can be concatenated for sequencing.

Tags from 5'end can be particularly reach in G/C, and their sequencing may require the use of additives in the sequencing reaction like betaine, which decrease the melting point of such structures (Shnoor et al. 2004).

On contrary to SAGE, CAGE is not restricted to the presence of an anchoring recognition site but only depends from the selection of true 5'ends and consequently from the capacity to synthetize and select full-length cDNAs in an effective way (Das et al. 2001).

In optimal conditions, using high-quality RNA, more than 90% of the full-length cDNAs are recovered, as observed from mapping the tags on full-length coding regions (Sugahara et al. 2001).

In contrast to SAGE for which a virtual reference database can be use for direct comparison, the analysis of 5'end tags entirely depends on strategies to map the tags to genomic sequences.

Full-length cDNA sequences and other sequence informations from transcribed regions can be used, although these are incomplete and nonrepresentative datasets, due to their low transcriptome coverage. Reference Sequences (RefSeq), which is a curated mRNA sequences collection, can integrate multiple forms of mRNAs into most likely transcript models, but often presents a larger version of an mRNA, because it contains more extensive 5' and 3' ends than other GenBank entries and the literature (Pruitt et al. 2005).

This has to be accounted in the interpretation of the CAGE tags mapping results.

In the case of CAGE tags from mouse and human tissues, mapping rates of 65-70% can be achieved using BLAST alignment (Altshul et al. 1997) and stringent mapping conditions.

Recently, CAGE data are available in a format which can be easily loaded into a Genome Browser (UCSC or Fantom3 Genome Browsers). Furthermore, the DNA Data Bank of Japan (DDBJ) has designed a new data format for deposing CAGE tags into public databases.

Whereas SAGE tags have defined ends which invariantly depend from the position of the anchoring enzyme recognition site in the genome, CAGE tags associated to the same TSS tend to cluster over short regions, displaying a slight variation in the position, owing to the inherent variability of TSS. Therefore, CAGE tags are grouped into "tag clusters" (CTSS or TK) comprising CAGE tags that map to the same genomic region and have at least one base pair in common.

One gene or transcriptional unit (TU) can have associated more than one tag cluster bacause it may have many different alternative promoters (Okazaki et al. 2002; Carninci et al. 2005).

Therefore, the information associated to one CAGE tag cluster is different from the use of a virtual tag per transcript in the annotation of SAGE tags.

In the case of nonclassifiable 5'-end tags, these can be confirmed by RACE experiments to ensure that they identify true TSS and to relate them to informations about transcripts, as it has been observed in a more systematic way for many of them (Kodzius R. unpublished observations).

The cross-validation by independent experimental means is particularly needed in the case of low-frequency tags derived from low-expressed transcripts. Some interesting considerations have been made comparing the results of two studies, which applied two different technologies, genomic tiling arrays and CAGE, to analyze the same type of human cell line (HepG2 liver cells).

In an initial tiling-tag comparison performed by sequencing 997,000 CAGE tags from total RNA of HepG2 cell line, only 20% of the tiling array transfrags had hits corresponding to CAGE tags. This was indeed expected, since CAGE tags are mostly concentrated around 5'ends of mRNAs, while the tiling transfrags are positive also in internal exons (Kapranov et al. 2002; Carninci et al. 2005). Futhermore, there was a dramatic decrease in the percentage of common hits when "novel transfrags" mapping 1kbp away from the current RefSeq annotations were examined, with only 2-3% of the CAGE tags matching the same positions. If the transfrags matching RefSeq are clustered together, more than 66% of these groups have overlapping CAGE tags within a 100bp distance, and even more 75% of the RefSeq clustered transfrags have CAGE hits within 1kbp distance. The dramatic difference in the overlapping between RefSeq-matched and novel transfrags suggest that novel transfrags detect rare RNAs that are expressed at less than one copy per milion, which could be more difficult to detect by CAGE with the current depth of sequencing (around 1milion tags). Alternatively, the novel transfrags may represent false positives in tiling arrays, or false negative hits, or some transcripts lack the cap structure (Carninci 2006). Furthermore, another possibility is that potential hits were discarded since they did not map to a unique genomic site (multimapped CAGE tags). When the partially overlapping CAGE tags were clustered together and compared to high-resolution 5bp-spaced tiling array transfrags from another study (Cheng et al. 2005) a

remarkable correspondence was found: 74% and 92% of the HepG2 CAGE tags map within 100bp and 1,000 bp distance respectively. This indicates that CAGE tag clusters reliably indicate the starting sites of mRNA identified by the transfrags and help to refine the 5'end structure of these transcripts. Out of 997,000 HepG2 CAGE tags, 97,353 represent distinct TSS, and are associated to 24,423 distinct transcriptional units, with an average of 3,98 tag clusters per TU. Among these CTSSs, 65,501 map within 1kbp upstream of a known mRNA or 5'EST, of which 41,974 are represented by only one tag, 7,444 have associated two tags whereas 16,083 have three or more hits.

Hence, most the tag clusters are represented by only one tag, suggesting that most of the transcriptional events in the cell are rare. This is confirmed by other studies, according to which in a HeLa cell, on average, there are about 3.5×10^6 copies of 28S 18S rRNA and together make up to 75% of cellular RNA, in the cytoplasm tRNAs and mRNAs comprise 10% and 2.5% of cellular RNA, with about 3×10^7 and 0.4×10^6 molecules respectively. Of these 400,000 mRNA molecules, a small number have many thousands copies per cell, while the vast majority have less than 10 copies (Bishop et al. 1974; Holland, Mayrand and Pederson 1980). Similar estimations have been shown for hepatocytes (Coupar, Davies and Chesterton 1978).

One of main advantage of CAGE is that the tags extracted by this technique are directly related to regulatory regions and therefore provide useful informations about the mechanisms behind the regulation of transcription in different cell context.

CAGE 5'end-tags together with 5'SAGE tags could then be used to build relationships between transcription levels, regulatory elements and trascription factors, and could potentially be related to other features of the transcriptomes, making feasible their application to gene network analysis and studies in systems biology (Ideker, Galitski and Hood 2001; Hieronymus and Silver 2004) where integration between experimental data and computer-aided theories requires that a significant amount of data are produced in an high-throughput way.

This latter point has still to be improved in the case of CAGE, because despite the great number of CAGE libraries have been analyzed (more than 400 mouse libraries with 7.1 milion CAGE tags and 236,000 TSSs and 40 human libraries with 5.1 milion CAGE tags and 180,000 TSSs), the classical CAGE protocol for tag production is still not easily applicable to an high-throughput scale.

Technologies to unveil transcriptome complexity: GIS, GSC, STAGE, DACS and other methods

The present sequencing-based approaches in gene-discovery are mainly limited by the end sequencing costs of randomly isolated clones. The discovery rate of different projects indicate that these approaches have not yet reached a plateau in their possible outcome (Carninci et al. 2003; Sheetz et al. 2004). Therefore, they remain still valid for driving the discovery and annotation of new transcriptional units and regulatory elements, as recently pointed out by deep sequencing of SAGE libraries (Pleasance, Marra and Jones 2003).

Recently, some more specialized methods have been developed. Among them, the gene identification signature (GIS) technique, where tags are sequenced from both ends of a cDNA (Ng et al. 2005).

5`end and 3`end tags derived from the same cDNA are combined into paired end-ditags (PETs) and cloned into concatamers for sequencing.

In GIS, a modified oligodT bearing the sequence for a methylation sensitive restriction enzyme (GsuI) is used for directional priming in the cDNA synthesis step. The cap-trapper method is employed to selectively recover only the full-length cDNA molecules and the digestion by GsuI is used to remove the terminal polyadenylated tract in the 3`ends, making them available for the ligation of an adaptor. Linker I and linker II contain the recognition site for a classIIs restriction endonuclease (tipically MmeI) and are ligated to the 5` and 3`endsof cDNAs. Ditags are generated

by the digestion with the classIIs enzyme and ligated to form concatamers for sequencing (Ng et al 2005).

The GIS approach overcomes the limitations deriving from the analysis of tags produced by two separate 5`end and 3`end tag libraries, improving the success rate in mapping the tags to the same transcriptional unit. In that sense, GIS could be seen as a better method for gene discovery and genome annotation, and could integrate 3`end SAGE-derived tags with 5`end tags obtained by CAGE or 5`SAGE.

However, a series of disadvantages prevent GIS to be applied to a very wide scale: (1) GIS libraries require special vectors for the library construction because commonly used standard vectors contain many MmeI sites in their sequence, (2) mapping rates for the half site (18bp) of a ditag is lower than for 20bp CAGE or 5`SAGE tags, (3) the identification of *trans*-splicing events may be difficult since both half sites of ditags have to map on the same chromosome (4) the larger length of the ditags has an impact on the quality and the cost of the sequences, (5) by introducing a bias against long fragments altering the representation of different transcripts. Last but not least (6) GIS cannot be applied to the analysis of poly(A)- RNA since using oligodT is mandatory for full-length cDNA synthesis from which both 5`ends  and 3`ends have to be recovered.

An alternative to the GIS method is represented by cloning the full-length cDNA into special vectors having classIIs recognition sites adjacent to the cloning sites, to avoid the need for special adaptors in the library preparation.

Other more specialized tag-based methods have been developed to analyze particular aspects of the transcriptomes and to interrogate and identify certain regions in the genome.

A method called digital karyotyping is based on sequencing tags derived from specific genomic loci defined by digestion with a "mapping enzyme". It aims to characterize large chromosomal changes associated to specific genomic regions of interest. It may be used to study naturally occurring copy number variations, which may have important functional roles associated to diseases.

Another approach is represented by DACS (digital analysis of chromatin structure), a quantitative method for a high-throughput mapping of DNAse I hypersensitive sites associated with *cis*-regulatory elements (Sabo et al 2004).

Another interesting method to characterize *in vivo* regulatory elements is STAGE, sequence tag analysis of genomic enrichment, which identifies direct binding sites for transcription factors on a genome wide scale (Kim et al. 2005, Bhinge et al. 2007).

Technologies to unveil transcriptome complexity: High-throughput sequencing technologies

DNA sequencing had a great impact on biomedical research and markedly influenced the development of genomic science. Large scale sequencing projects, including whole-genome sequencing, have usually required the cloning of DNA fragments into bacterial vectors, amplification and purification of individual templates, followed by Sanger sequencing (Sanger Nicklen and Coulson 1977) using fluorescent chain-terminating nucleotide analogues and either slab gel or capillary electrophoresis (Prober et al. 1987).

Current estimates evaluated the cost of the human genome in the 10 and 25 milion USD range (http://wwwgenome.gov/12513210). Alternative sequencing methods have been described (Bains and Smith 1988; Jett et al. 1989; Jacobson et al. 1991). However, no other technology has overcame the need for cloning in bacterial vectors while Sanger sequencing remained the most widely used method for generating DNA sequence data.

More recently, a great effort has been made to find alternative and more cost-effective strategies for DNA sequencing, which will have a significant scientific, economic and cultural impact.

In particular, the so called pyrosequencing has started some time ago to emerge as a good alternative method for sequencing short DNA fragments. This technique enables sequencing of millions of bases, by integrating an emulsion-based method for isolating and amplifying DNA

fragments in *vitro*, and an instrument which performs pyrophosphate-based sequencing in picoliter-sized wells (Margulies et al. 2005).

Pyrosequencing is a method of DNA sequencing based on the "sequencing by synthesis" principle developed by Mostafa Ronaghi and Pål Nyrén (Ronaghi et al. 1996; Ronaghi, Uhlen and Nyrén 1998; Nordström T et al. 2000; Nyrén 2007). Pyrosequencing AB was started to commercialize the machine and reagents for sequencing of short stretches of DNA. Then, it was renamed to Biotage and licensed to 454 Life Sciences. 454 developed an array-based pyrosequencing which has emerged as a rapid platform for large-scale DNA sequencing. Latest platform of pyrosequencing (GS FLX from 454 Life Sciences which is owned by Roche) can generate 100 million nucleotide data in a 7 hours run with a single machine. It is anticipated that the throughput may increase by 5-10 fold with the next release. Since each run may cost about 8 to 9 thousand USD, *de novo* sequencing of mammalian genomes are in million dollar range. Therefore, pyrosequencing opens the way to many applications to be achieved in a shorter time and at a fraction of the cost of the Sanger sequencing.

More recently, other platforms for high-throughput sequencing became available: Solexa sequencing by synthesis (Illumina, Inc., San Diego, CA, USA), SOLiD sequencing by oligonucleotide ligation and detection with 2-base encoding (Applied Biosystems, Foster City, CA, USA), and true single molecule sequencing by Helicos (Helicos BioSciences Corp., Cambridge, MA, USA).

Each of these platforms have their own advantages but all of them skip the concatenation step needed by the 454 system. In this way the number of bp that can be read in 1 day by high-throughput sequencers is orders of magnitude larger than conventional sequencers based on Sanger sequencing (see Table 1).

These features makes them particularly suitable for short sequence reads of the kind generated by CAGE, SAGE and other tagging technologies, in contrast to genome sequencing, for which longer

sequence reads are preferred. Hence, transcriptome sequencing and SNP genotyping projects can be primary applications of high-throughput sequencers.

Table 1. Sequencing Capacities of Current and Near-future Next-generation Sequencers, and the Estimated Corresponding Deep CAGE Throughput and Sensitivity

Sequencer	Throughput				CAGE	
	Base/Run	Base/Read	Read/Run	Base/Day	Mapped Tag/Run	Sensitivity at 99%
454	100,000,000	250	400,000	100,000,000	1,100,000	1 per 1 cell
Solexa	500,000,000	25	20,000,000	120,000,000	5,000,000	1 per 5 cells
SOLiD	1,400,000,000	35	40,000,000	180,000,000	13,000,000	1 per 14 cells
Helicos	10,000,000,000	30	300,000,000	1,000,000,000	150,000,000	1 per 160 cells

The sensitivity of deep CAGE is assessed as the concentration per cell of an mRNA molecule that can be detected at 99% probability, assuming a total mRNA concentration of 200,000 molecules per cell.

taken from De Hoon and Hayashizaki 2008

The increased throughput of sequencers enables CAGE sequencing at a much deeper scale than previously possible. This implies that it is becoming increasingly possible to detect RNA molecules present at very low-copy numbers in the cell. This could be critical in time-course experiments, for instance, where transcripts coding for regulatory molecules may be present only transiently and at low concentrations.

But how many tags would be enough to sequence in order to have the maximum depth of coverage of all the transcripts present in a given cell at a given time?

Approximately, the probability to detect an RNA molecule that is present in k copies per cell is equal to $1 - \exp(-nk/m)$, where n is the number of CAGE tags extracted and m is the total number of RNA molecules in the cell. A 454 sequencer (454 Life Sciences, Branford, CT, USA) generating 400,000 reads per run of a length of about 250 bases can produce ~1,000,000 mappable CAGE tags in one run. Assuming a total mRNA concentration of 200,000 molecules per cell, the probability to detect an RNA molecule present in one copy per cell is 99.33%. The Solexa sequencer can generate about 5× more CAGE tags per run, making it possible the detection of an RNA molecule present in one copy per five cells (De Hoon and Hayashizaki 2008).

Recently, Zhu and colleagues found that sequencing tens of thousands of tags in an ordinary EST/SAGE experiment was far from sufficient to cover all the transcriptome of single cell types.

Technologies to unveil transcriptome complexity: PCR-based and Transcription-based methods for RNA amplification

Over the past few years a growing need to bring the gene expression analysis to a scale of few cells has contributed to develop a variety of different approaches to amplify the RNA from amount equivalent to single or few cells. It has been estimated that a single eukaryotic cell contains approximately 0.1pg of mRNA and 10pg of total RNA. This amount would need to be amplified $10^6 - 10^8$ fold to generate enough DNA/RNA targets for hybridization on two-channel microarrays. In order to solve this problem and overcome the technical inherent limitations of the current array technology, a series of different strategies have been explored, which can be basically divided into two broad categories: signal amplification and sample amplification.

In the first strategy the main goal is to improve and optimize the labelling reaction increasing the number of signal molecules per transcript; this can be achieved using technologies such as dendrimer (Stears, Getts and Gullans 2000), or tyramide signal amplification (TSA) (Karsten et al 2002). The dendrimer application uses two-step hybridization, in which the cDNA is synthesized with an olidodT primer containing a capture sequence and then hybridized to the microarray slide. In a subsequent step, dendrimers each containing multiple fluorescent molecules, are hybridized to the capture sequence. Genisphere claims that 0.25-1 µg of total RNA ($2.5 \times 10^4 - 10^5$ cells) can be used as starting material for having a good quality array preparation using dendrimer technology.

For the other strategy, focused on sample amplification, two different approaches have been used: methods which make use of exponential polymerase chain reaction (PCR) and methods which employ linear isothermal amplification by *in vitro* transcription (IVT) of the cDNA into labeled complementary RNA (cRNA) (reviewed in Nygaard and Hovig 2006).

All of them present both advantages and drawbacks, but all of them have some intrinsic bias, which are also function of the starting amount of RNA (Zhao et al 2002), of whether antisense or sense RNA is produced (Goff et al 2004) and of the number of rounds of amplification performed. In addition, time-dependent RNA degradation during IVT can introduce some noise into the resulting microarray data (Spiess, Mueller and Ivell 2003).

The methods which make use of linear isothermal amplification by *in vitro* transcription (IVT) have been originally developed by Van Gelder et al. in 1990, commonly known as Eberwine method.

This method consists of a general RT step using an oligodT primer, bearing a T7 RNA polymerase promoter sequence, which is used to *in vitro* transcribe the cDNA into antisense RNA (aRNA) after conversion of the mRNA/cDNA hybrid into ds-cDNA (Van Gelder et al 1990). This T7 IVT-based method has been used in several studies, scaling down the RNA amount to single cell level (Eberwine et al 1992, Phillips and Eberwine 1996, Crino et al 1998, Chow et al 1998). aRNA has been radioactively labeled and hybridized to either known genes on dot blots or to cDNA libraries from specific tissues (Crino et al. 1996 among others).

Many later studies have been showing the technical limitations of the Eberwine method and have modified the procedure in order to optimize and improve the reprobucibility and to reduce some primer-dependent artifacts produced during *in vitro* tramscription in absence of a legitimate template. In particular two pioneer studies were the first to quantitatively evaluate the differences in gene expression profiles between amplified and unamplified samples, using cDNA or oligonucleotide arrays (Wang et al. 2000; Baugh et al. 2001).

Baugh and collaborators have maintained almost the same protocol used by Eberwine but they noticed that lowering the volume of RT reaction and the amount of RT primer could avoid the accumulation of primer-dependent artifacts in the subsequent IVT step. Furthermore they have shown that using single stranded binding proteins like T4gp32 could improve the processivity of the RT reaction *in vitro* as previously reported for other retro-viral capsid proteins (Ji, Klarmann and Preston 1996).

On the other hand, Wang and colleagues have used a template switching effect at the 5'end to ensure a full-length cDNA synthesis and solve, at least in part, the problem of the 3'bias in the original Eberwine method. The template switching effect is based on the addition of non template residues, primarily cytosines, to the 3' end of the cDNA, in correspondence of the cap structure, by the reverse transcriptase, with a mechanism not completely understood. The use of an oligonucleotide containing a 3'end polyG tract in the RT reaction mixture, which pairs with the newly synthesized dCTP stretch, allowes the reverse transcriptase to switch template and extend the annealed oligonucleotide, producing full-length cDNAs (Wang et al. 2000). A subsequent IVT step has been used to synthesize aRNA, which has been then labeled and hybridized to cDNA microarrays (Onco-chip containig 2,008 human genes). Both studies (Wang et al. 2000; Baugh et al. 2001) concluded that the concordance between amplified and unamplified samples was high (R^2 0.99-0.96) although some discrepancies emerged, increasing as the amount of starting total RNA decreased.

In particular Wang et al. found that a labeling bias affected low transcript levels in which the background of the fluorescence in one channel was higher than the signal, but not in the other channel, resulting in an increased false positive rate when the amplification started from amount of total RNA comprised in a range of 0.125-0.001 μg, and that this bias is in part recovered if a second round of amplification is introduced ensuring 3 μg, usually considered as lowest limit for the target for array hybridizations. A series of other studies followed these pioneer works introducing some modifications to improve the fidelity, sensitivity or reproducibility of the T7 IVT-based method (Zhao et al. 2002; Moll, Duschl and Richter 2004; Naderi et al. 2004; Xiang et al. 2003).

As alternative to the T7-based linear amplification methods, PCR-based exponential amplification strategies were introduced. These methods share the use of a global PCR step which is made possible by introducing PCR-priming sites at both ends of each reverse transcribed cDNA molecule (Hertzberg et al. 2001; Iscove et al. 2002; Li et al. 2003).

Li and collaborators applied the SMART[R] technology, based on the same template switching effect previously described, to introduce the PCR-priming site at the 5'end and to ensure full-length cDNA at the same time. In another approach the template switching-generated ds cDNA was used to amplify selectively only one of the two strands then used in multiple rounds of reverse and *in vitro* transcription (Stirewalt et al. 2004).

The PCR-based methods present several advantages over linear amplification strategies, as they usually yield better amplification rates ($1\text{-}3\text{x}10^{11}$) starting from equivalent amounts of total RNA as template, are less labor intensive and consequently also more cost effective, and the ds-cDNA products generated are suitable for hybridization to array probes of either orientation, and are also more stable than the aRNA products (Wang 2005; Nygaard and Hovig 2006).

At the same time, PCR-based methods have also a number of inherent limitations: first of all the relevant properties of the DNA polymerases used in the reaction such as misincorporation of bases, bias toward shorter transcripts and differential amplification efficiencies of templates which have different GC composition are important factors which can affect the final output of the protocol.

The two latter properties result in non linear amplification, which could lead to a misrepresentation of the quantitative transcript values in a sample after performing multiple PCR cycles. Use of the DNA polymerases could achieve a more linear representation of the transcript levels when using single primer amplification (SPA) as in the method described by Smith et al. 2003. In this case the primer extension actvity of the DNA polymerase, similar to a cycle sequencing reaction, has been employed instead of the double primer exponential polymerase chain reaction. A modified oligodT bearing a heel sequence (derived from the T7 promoter) was used for priming the reverse transcription and the ds-cDNA product was denatured and annealed to a single primer complementary to the heel sequence, and then extended in multiple cycles by the *Taq* DNA polymerase (Smith et al. 2003).

The Pearson correlation coefficients for two self-self hybridizations derived from two independent SPA amplifications of both kidney and liver cDNA were 0.985 and 0.986 respectively, demonstrating a semi-linearity with values comparable to those of T7-based methods.

A subsequent improvement of this strategy led to a linear isothermal amplification of cDNA using a single primer (Dafforn et al. 2004).

Another issue is the strand-specific amplification: this has no implication when cDNA microarrays are employed but becomes important when oligonucleotide chips are used, since hybridization to oligonucleotide arrays requires the antisense strand of the nucleic acids.

Unlike the PCR-based strategies which generate ds-cDNA, the *in vitro* transcription-based methods can produce either sense or antisense RNA, depending from the protocol used.

The strategy adopted by Affymetrix involves the synthesis of biotin-labeled complementary RNA (cRNA) which can be directly used for hybridization.

However, the use of cRNA for target preparation presents some potential disadvantages, like the limited stability of the target and the lower specificity of RNA-DNA interactions compared to DNA-DNA, due to the increased stability of binding energy for the mixed hybrids, which are thus less sensitive to mismatches.

In the case in which a template switching mechanism is applied with a primer matching the 3'end of the first strand to generate the ds-cDNA, the subsequent *in vitro* transcription sense RNA can be retrotranscribed into labeled antisense cDNA target suitable for oligo-array hybridization.

Alternatively, the antisense RNA, normally generated by IVT-based protocols, can be retrotranscribed by 9-mer random primers bearing a T3 RNA polymerase promoter, which result in a sense-stranded RNA (Kaposi-Novak et al. 2004; Che and Ginsberg 2004).

Irrespective of whether a linear or exponential method is employed, the applicability of a sample amplification strategy is evaluated by estimating a series of common aspects which can be summarized as amplification efficiency, reproducibility, 3' bias and product length and fidelity in maintaining the relative transcripts abundance.

The amplification efficiency is defined as the ratio between the final aRNA product and the initial mRNA imput or the cDNA yield and the initial mRNA imput in the case of PCR-based methods. In general the total RNA content is measured and the mRNA is assumed to be a fraction variable from 1 to 3%, depending on the type and the metabolic activity of the cells.

Besides technical issues related to the particular amplification protocol of choice, the amplification fold factor is affected by RNA measurements and theoretically estimated mRNA content.

For linear IVT-based protocols, a 10^3-10^5 fold amplification factor has been reported after two rounds of *in vitro* transcription (Baugh et al. 2001), whether for PCR-based amplification an average of 10^{11} fold factor is normally achieved (Iscove et al. 2002).

Whereas for methods which use a mixed strategy conjugating both linear and exponential steps, a 10^6-10^7 fold amplification factor has been reported, which lies between that of two rounds of linear amplification and exponential amplification.

Most of the amplification methods, described till now, make use of an oligodT primer for the first round of cDNA synthesis and a random primer is usually used for the second round of amplification, leading to shortened products. The transcript lengths are dependent on the processivity of the reverse transcriptases and polymerases used and on the reaction conditions.

After two rounds of amplification, product lengths, as assessed by gel electrophoresis or Agilent Bioanalyzer, are tipically in the range of 200-1000 bp with the peak at 400-500 bp, with a slight variation depending from the particular technique applied. In most of the cases, the amplified targets present a 3'end bias due to the shortening effect concomitant with the 1st to 2nd round transition. As observed by Baugh et al. the signal obtained from oligo probes which are 5'bias is markedly reduced due to the 3' target bias and the lack of full-length products. However, this 3'bias has only a minimal effect on the oligoarray and cDNA microarray analysis of gene expression levels, considering that generally most of the probesets also exhibit a similar 3'bias.

Regarding the reproducibility, many studies have shown that most of the amplification methods ensure on average highly reproducible results. Parallel but independent amplifications yield highly

correlated expression profiles (Wang et al. 2000; Baugh et al. 2001), even if diluting the imput RNA down to tens of nanograms (r 0.94). However, parallel studies have tried to assess whether serial dilutions of the starting RNA could affect the reproducibility rate, and actually the variabilty in the results increase as the quantity of RNA amplified was reduced (Kenzelmann et al. 2004).

A comparison of the linear versus exponential amplifications, also revealed that the results produced from these two strategies are slightly different, although some of these comparisons have shown discordant results: in one study linear amplifications have been shown to be slightly more reproducible than exponential methods (Puskas et al 2002), while a second study found an opposite trend (Klur et al. 2004).

Another important aspect that has to be considered is the assessment of the fidelity of the method of choice: this can be defined as the extent at which the abundance of the different transcripts is faithfully preserved. It's evident that fidelity is a prerequisite that has to be satisfied to apply the amplification technique in combination with microarray for accurately and quantitatively profile gene expression levels. Calculation of Pearson correlation coefficients has been widely used as a common statistical approach when comparing different expression profiles.

Evaluation of consinstency of the outliers between non amplified and amplified samples has also been frequently used as parameter to assess the fidelity of the ampification procedure. In particular Wang et al. based all their analysis only on the mantenance of differentially expressed genes between two different RNA sources, revealing that when 0.25-3 μg of total RNA were used for IVT-based amplification, 85-92% of the outliers from the control unamplified sample were also found to be differentially expressed above the threshold (ratio greater than 2-fold) in the amplified dataset. Another approach has been used to show all the genes in common between non amplified and amplified samples, and although the absolute intensity levels were different, the relative abundance were preserved, with a relatively high correlation factor (r 0.84) (Scheidl et al. 2002).

However, for those genes whose signal intensity approach the background level on the array, the correlation coefficients dropped down. In that case most of these discordant ouliers could be recovered as another round of amplification has been performed as observed from Wang et al. 2000. Even more informative statistical analyses were conducted focusing on the components of a variance model (Zhao et al. 2002). In this study the authors have examined the variance of true expression and measurement errors for both array signals amplified with different methods and unamplified array signals and have found that a decreased variance in gene expression occurs after amplification, suggesting a dampening effect of the amplification itself on the true gene expression values. A compression effect, at least in the case of SMART-based amplification methods, as consequence of the bias introduced by PCR, have been also documented by other independent studies (Iscove et al. 2002; Petalidis et al. 2003). In particular Petalidis et al. observed that log ratio values of the genes identified as being differentially expressed between samples decreased with increasing numbers of PCR cycle amplification. This seems to suggest that PCR-based methods tend to dampen the relative ratio of the more highly differentially expressed transcripts, as well as raise up the level of the low expressed transcrips. Since they exhibit a level of expression close to the background in the non amplified (direct labeled) sample, their relative ratio after amplification is increased making appearing them as a group of differentially expressed genes while they are not. This is likely to be a consequence of the non-linearity of the PCR amplification process. The increase of the number of PCR cycles will raise the level of low represented transcripts, bringing the most highly expressed genes to reach a plateu, thus flattening the expression ratios for most of them. This incovenient could be presumably overcome assessing the most appropriate number of cycles to be used in order to preserve the relative expression levels of the largest part of the genes and maximize the fidelity of the procedure.

Petalidis and colleagues have been used a simple template-switching PCR protocol for the amplification. By comparing the amplified samples to directly labelled, non amplified samples, they have found that 19 cycles of PCR exhibited the best performance in terms of fidelity and accuracy,

having the maximum true outlier identification rate at the expenses of a slight decrease in the total number of genes above the background. Surprisingly, even a modest increase in the number of cycles above 19 have been shown to severly affect the number of true outliers.

The reported results from these analyses have been restricted to outliers or genes with a fold change of more than 2. However recently the 2-fold change criterion as a measure of significance has been replaced by the application of t-tests for sub-sets of genes, and, in addition, genes co-regulated in pathways or signatures rather than single differentially expressed genes have been extensively used. A comparison of t-scores or distribution of fold change for individual genes have been applied (Gold et al. 2004; Li et al. 2005; Nygaard et al. 2005), usually restricted to a small subset of genes, like the top ten ranked outliers.

A general conclusion comes out comparing all of the more recent studies: the number of differentially expressed genes increases when amplifying from decreasing amounts of imput RNA.

In a recent study (Nygaard et al. 2005) it has been found that only moderate/highly expressed genes were quantitatively reliable in experiments starting from less than 1000 cells, whereas low expressed genes were subjected to stochastic fluctuations, which limited the precision of gene expression measurements.

CAGE data and the analysis of mammalian promoters

CAGE data and the analysis of mammalian promoters: The classical concept of "core promoter"

Many aspects of the cellular activities in eukayotes like homeostasis, growth, differentiation and development rely on the regulated production of specific mRNAs by RNA polymerase II (RNApolII). The mechanisms that underlie this regulation have been the subject of intense genetic, biochemical and computational studies (Smale and Kadonaga 2003; Gross and Oelgeschlager

2006). The transcription start site (TSS) of a gene is defined as the first nucleotide which is copied at the 5'end of the corresponding mRNA.

The region around a TSS is often referred to as the "core promoter", which is required for recruitment of the transcription apparatus and can be thought of as the priming stage site for transcription initiation. Due to the strong link between TSSs and core promoters, the terms are often used interchangeably. In the case of many genes this can leads to a misunderstanding, since most of the genes can have more than one TSS and transcription starts from slighlty different positions. In these cases, the core promoter is defined like the region that spans all the different TSSs.

The classical textbook model of an RNA polII core promoter is represented by an AT-rich motif (TATA-box) approximately located 30bp upstream of an Initiator (Inr) element which contains the TSS. Two different models for the assembly of the pre-initiation complex (PIC) have been described for TATA-box bearing core promoters: in both models the binding of the basal transcription factor TFIID to the TATA-box and the Inr sequences (-35 to -19bp from TSS) which causes a bending of 90° of the DNA, is the first event required for the formation of a stable basal transcription machinery (Burley and Roeder 1996). The TFIID binding is facilitated by TFIIA which seems to prevent the inhibitory effects of other factors such as Dr1 and Dr2 (Drapkin et al 1993). Then in the case of the so-called stepwise assembly model, the progressive association of other basal factors TFIIB-H and of RNApolII leads to the formation of a competent RNA polymerase II holoenzyme, which with other factors as the Mediator Complex, co-activators and co-repressors coordinates both basal and regulated transcription initiation (Lewis and Reinberg 2003; Black et al. 2006; Thomas and Chiang 2006).

Coordination of chromatin modification, mainly through the control of post-translational modification of histones, also plays an important role in transcription initiation (Black et al. 2006). Furthermore,,the recruitment of all of these co-activators and co-repressors of transcription initiation is controlled by transcription factor binding to *cis*-acting DNA sequences that can lie

within the core promoter or in more remote locations (enhancers and repressors) (Wasserman and Sandelin 2004).

Apart from the TATA-box, subsets of promoters contain the Inr element, CpG islands and other sequence patterns, but their prevalence and role in the initiation of transcription are not as well characterized.

In summary, several common DNA elements have been characterized in core promoters, and are linked to the expression of the downstream genes. Different elements can co-occur in the same promoter, although certain combinations are more likely than others.

The TATA box is located 28-34 bp upstream of the TSS. It has the consensus TATAA. This sequence is bound by the TATA-box binding protein (TBP) which recruits the other components of the pre-initiation complex (PIC).

From the inspection of large CAGE datasets became evident that TATA-box are mainly associated to strong-tissue-specific promoters and often co-occur with Initiator-like motifs.

The Initiator element is defined by the consensus YYANWYY where A represents the +1 position from the main TSS. This DNA element can function independently from the TATA-box although these two elements often act sinergistically in the same promoters. The Initiator element alone is able to recruit the PIC complex (Smale and Kadonaga 2003).

A class of TATA-less promoters in *Drosophila melanogaster* was found to share a common motif 28-32 bp downstream of the TSS, called downstream promoter element (DPE), (Ohler etal 2002, Kadonaga JT 2002). It has the consensus RGWYV and generally is found together with Initiator elements and functions similarly to the TATA-box, recruiting the PIC complex to a close TSS.

Another common element is the TFIIB recognition element (BRE) which has the consensus SSRCGCC and is located upstream of the TATA-box in some TATA-dependent promoters. It may contribute to modulate the strenght of the promoter tuning the rate of transcription in an unknown way (Lagrange etal 1998).

Another DNA sequence pattern often associated to eucaryotic promoters is represented by CpG islands. These are genomic stretches where CG dinucleotides are overrepresented. Based on the most recent statistical definitions, 72% of human promoters seem to be associated to a CpG island (Saxonov etal 2006). CpG islands are most often associated to housekeeping or ubiquitously expressed genes, but there are also many exceptions like for instance brain-specific genes (Gustincich etal 2006).

CAGE data and the analysis of mammalian promoters: classification of promoters classes

The CAGE technique has disclosed several unexpected features of the core promoters associated to specific classes of TSSs paving the way for a more reliable description of the transcriptional regulation mechanisms. The most extensive core promoter identification study undertaken so far used CAGE tags to identify 184,379 human and 177,349 mouse core promoters, many of them being associated to at least one tag cluster (Carninci etal 2006). This genome-wide analysis indicated that most human and mouse promoters lack the distinct TSS commonly assumed to be located at one precise genomic position along the different TUs. Instead, the typical promoter architecture consist of an array of closely located TSSs spreading in an interval of 50-100 bp. These observations give the opportunity to employ a new system for promoter classification.

Breafly two main categories can be distinguished, the "sharp" core promoters with single TSS clusters and the "broad" core promoters identified by broad TSS regions. Among these two categories there are some hybrid cases.

In a more detailed study (Carninci etal 2006), 8,185 mouse and 5,928 human tag clusters supported by at least 100 CAGE tags were used to classify TSS distributions into four "shape" categories. The single dominant peak class (SP) where the majority of tags are concentrated around a single dominant TSS position, and other three broad cluster distributions: a general broad distribution (BR), a broad distribution with a dominant peak (PB) and a bi- or multimodal distribution (MU) in

which several main TSS can be distinguished.

These four classes identify different promoter context. Using position-specific weight metrices, TATA-boxes seem to be overrepresented in promoters with sharp TSS, whereas CpG islands are associated to broad TSS regions. In 90% of the cases TATA-independent transcription initiation occurs in CpG islands, and this data was quite unexpected based on the previous studies when researchers mainly focused on studying highly expressed and tissue-specific genes, which usually contain a TATA-box in their core promoter.

CCAAT-box and GC-box that are not associated to CpG islands are preferentially associated with Single Peak –type TSS.

Another important feature emerged from the statistical analysis of the CAGE data was the estimation of the frequency of the initiator element and its consensus.

58,6% of the CAGE tags have a pyrimidine-purine dinucleotide at position -1/+1. This pyrimidine-purine dinucleotide corresponds in part to the Inr element or Cap motif which was determined by mutagenesis studies (Smale and Kadonaga 2003). Sequence logos for promoters aligned to TSS clusters have been constructed counting each tag and its flanking region as one sequence divided by promoter shape classes. This analysis unveiled that the most preferred initiators are CG, CA and TG more often associated to highly used TSS, whereas GG initiators are preferentially associated to low expressed transcripts (Carninci etal 2006).

Substitutions of the pyrimidine-purine dinucleotides correlate with large differences in TSS usage. Transversion-type of substitutions negatively affect the rate of initiation whereas mutations from other sequences to pyrimidine-purine dinucleotides create new TSSs. Comparison of the mouse and human broad promoters confirm the general evolutionary conservation of TSS position inside the clusters.

More recently, it has become clear that transcription initiation events are clustered on the cromosomes at multiple scales, with clusters within clusters, each subjected to multiple regulatory processes (Frith etal 2008). The smallest of these clusters is represented by core promoters and the

local sequence around TSS can predict the relative transcription start usage of each nucleotide position with a 91%-accuracy, implying the existence of a DNA code that determines TSS selection. Conversely the global transcriptional activity of a locus is controlled by distal features such as enhancers and chromatin state. Martin Frith et al. used a HepG2 CAGE dataset and applyed a Markov Model strategy to calculate the frequency of all k-mers combinations in a -50/+50 bp interval around each main TSS in the genome and predict the TSS usage rate directly from the local DNA sequence.

This analysis revealed that there are three main motifs overrepresented in the -50/+50 interval: the most conserved and overrepresented is the TSS itself, which confirmed the main role of the pyrimidine-purine dinucleotide in determining the position for initiation events. But taking into account only this information (-1/+1 dinucleotide), the rate of success in TSS position prediction is 78%. Therefore, other local sequence informations are necessary to correctly guide the transcription initiation events. The second most overrepresented k-mer is the TATA-box, although it is not present in the majority of the core promoters. The third overrepresented motif is a GCG trinucleotide repeat region, located downstream of the TSS. Actually, an Sp1-like binding motif is often found in the -50/-30 region which is thought to play a role similar to TATA-box in recruiting the PIC complex in TATA-less promoters (Butler and Kadonaga 2002).

This GCG repeat is often found to have an "echo" GCG repeat 10nt downstream of the TSS, which is located in the same phasing relative to the TSS, since 10nt is one turn of the double helix. The Inr, Sp1, gcg and gcg echo motif tend to co-occur with each other more often than expected by chance (Frith et al 2008).

CAGE data and the analysis of mammalian promoters: transcription starts from unconventional sites (Exons, introns and 3`UTR)

Not all CAGE tags map to previously identified 5`ends of full-length cDNAs. Typically, some of the main CAGE tag clusters mapped to 5`end but some CAGE tags were also associated to the

3`UTR region. The CAGE tags that map to genomic regions between these two main peaks map mostly to exons. In the case of relatively highly expressed genes, the internal exonic tags are supported by a considerable number of 5`EST and have been confirmed using RACE, based on the oligo-capping method (Suzuki etal 1997). In the comparative study of promoter architecture based on 159,075 CAGE tag clusters in mouse, 34,229 TSSs (21,5%) mapped within exons and presumably would generate transcripts that truncate or eliminate the predicted protein product (Carninci etal 2006).

The exonic promoter activity varies among genes and it is often conserved across species.

Exonic promoter activity does not correlate with the number of tags of the major promoters, but in general correlates with tissue-specific genes having a single dominant peak of transcription initiation (sharp promoters). This is not reciprocally true since there are many single peak sharp promoters that do not display any internal exonic associated tags.

The evolutionary conservation of the internal exonic TSSs suggest that the transcripts produced from these sites are functionally important and they do not respresent just experimental artifacts, since they are confirmed also by other independent techniques like genome tiling arrays.

There is a good coverage of known RefSeq transcripts exons with both arrays and tags. Most RefSeq exons can be identified by tiling arrays and most TUs (>70%) can be identified by CAGE tags, despite largely incomplete overlap in the tissues examined with the two technologies (Carninci etal 2005).

However, CAGE tags detect both longer and shorter versions of annotated transcripts. Some of these transcripts partially overlap coding mRNAs (Carninci et al 2005), including the TSSs originating from within exons. These RNAs might not encode proteins; however, transcription is conserved between mouse and human, suggesting that these seemingly shorter transcripts, or transcription itself, can be functional.

Although these RNAs could only be speculated as being involved in splicing regulation (Bentley etal

2005), it is noteworthy that ChIP mainly detects the hypophosphorylated form of RNA Pol II, which is associated with elongation and with alternatively spliced exons but not with the introns of human genes (Brodsky et al 2005).

Brodsky *et al.* analysed the locations of active DNA-bound RNApolII in HeLa cells. Surprisingly, they found that RNApolII sites were concentrated preferentially in exons. The density of RNApolII sites in exons varied between genes, but did not correlate with mRNA levels. The authors attribute these unexpected results to a possible slowdown or pausing of RNApolII elongation within exons (Bai et al 2006). Indeed, an earlier study showed that variations in the speed of RNApolII elongation will affect the usage of splice sites (Kornblihtt et al 2004).

Accordingly, the amount of exonic RNApolII sites was greater in alternatively spliced exons compared with the invariantly spliced ones (Brodsky et al 2005). These observations suggest that truncated internally initiated mRNAs constitute a significant class of noncoding mRNAs; this is also consistent with evidence that the initiation complex can bind to sequences within exons but not introns (Kim TH et al 2005).

The possible function of weak initiation sites within internal exons requires further study, but it could contribute to the recently described phenomenon of exon-tethering — a physical connection between emergent splice sites in pre-mRNA and the RNApolII transcription complex (Dye et al 2006). ChIP experiments and the complexity of TSSs show that transcription factors not only bind to putative promoter regions but also frequently bind to regions within or downstream of annotated genes (Cawley etal 2004).

Another unconventional site in which many genes display promoters is represented by 3`UTR.

There is a considerable increase in CAGE tags incidence in the 3` UTRs of protein-coding transcripts. These TSSs have been independently validated by a 5`RACE based on oligo-capping and are also supported by GIS and GSC ditag analysis (Carninci et al 2005, Katayama et al 2005).

These TSSs also have a distinct sequence motif respect to tags in 5`end. Alignments of the most

tag-rich TSSs derived from 3`UTRs revealed a strong overrepresentation of three consecutive guanines (consensus RGGG), found at position –3 to–1 in 785 out of 1,327 cases, just before the TSSs. Analysis of cross-species conservation between sequenced vertebrate genomes in the region surrounding the 3`UTR TSSs revealed a highly conserved region located at positions +40 to +90 relative to the TSSs (Carninci et al 2006).

To confirm that these sequences can indeed initiate transcription, a reporter-gene assay has been performed with four distinct 3`UTR promoters. In each case, upstream regions of the TSS directed reporter-gene expression.

Transcripts initiated in 3`UTRs might regulate downstream genes using a sense-antisense mechanism, as downstream genes on the opposite strand are located much closer than expected (Katayama et al 2005). If 3`UTR–derived transcripts function as regulatory noncoding RNAs, their transcriptional regulation might be discordant from the full-length transcript.

In 43% (168/391) of testable representative transcripts, the tag distribution in 5` and 3` terminal exons was significantly divergent in at least one tissue, suggesting independent regulation of the 3`UTR promoter (Carninci et al 2006).

Notably, the incidence of 3` TSSs is tissue specific: it is prevalent among CAGE libraries derived from cerebellum and lung but reduced in libraries derived from embryo. Additional studies will be needed to further characterize a possibile mechanism of activity and functions of the 3`UTR promoters.

CAGE data and the analysis of mammalian promoters: Widespread occurrence of alternative promoters

Mammalian genomes make an extensive use of alternative splicing as a key mechanism to expand transcript and protein diversity. In this way different exon combinations are used to generate multiple mRNAs from a single gene to encode different protein isoforms with diverse and even antagonistic functions (Pajares et al 2007). The mRNA isoforms are usually transcribed by

alternative promoters including alternative first exons. Numerous genes displaying complex transcriptional regulation, because of the use of alternative promoters, have been identified (Ayoubi and Van de ven 1996, Landry 2003). Several genome-wide analyses indicate that >60% of the human genes use alternative splicing. Recent evidence suggests that 30–50% of the human and about 50% of the mouse genes have multiple alternative promoters that can span up to thousands of kilobases (Kimura et al 2006, Sun et al 2006, Baek et al 2007) . A comprehensive analyses of the ENCODE (http://genome.ucsc.edu/ENCODE/) regions (The ENCODE Project Consortium 2007), which corresponds to 1% of the human genome in 16 diverse human cell lines, using transient transfection reporter assays, identified >20% of genes having functional alternative promoters (Cooper et al 2006). One specific case is examplified by the erythroid genes of which 35% had evidence of alternative first exons and promoters in humans (Tan t al 2006). Kim et al. identified 10,567 active promoters corresponding to 6,763 known genes in human fibroblast cells, using high-density DNA microarrays that represented all the nonrepeat DNA throughout the human genome at 100-bp resolution. Their analyses revealed 1,609 genes (24%) with active multiple promoters. This is quite a high percentage of genes showing multiple promoter use in a single cell type, and the other inactive promoters might be active in different cell types or under different environmental conditions, suggesting extensive use of multiple promoters by mammalian genes (Kim et al 2005). More recently, a genome-wide computational analysis of mouse and human genes demonstrated a strong functional correlation between usage of alternative promoters and presence of alternative splicing. Interestingly, the number of alternative promoters used by a gene is positively correlated with the number of alternative splicing forms (Xin, Hu and Kong 2008).

These studies indicate the prevalence of alternative promoters in mammalian gene expression regulation but the functional significance and their role in disease etiology remains almost completely unexplored.

There are several ways to create diversity in time and space in regulatory control of gene expression. One is to provide a particular promoter region with all the *cis*-acting control elements

needed for the tissue- or cell-type- specific and/or developmental stage-specific expresssion and for its response to hormones or other signals. Even genes expressed in a single cell type, like albumin in hepatocytes, may have a very complex array of *cis*-acting elements to which a large number of *trans*-acting factors bind, ensuring the proper timing, level of expression and cell specificity of the gene (Crabtree Schliber Scott 1992).

Another way to generate diversity in expression is to duplicate a gene and to provide individual copies with their own regulatory elements. An example of this is the globin gene family for which the timing of expression of the various isoforms during embryonic development has to be titghly regulated.

Alternative promoter usage resulting in alternative leader exons can affect the gene expression of the corresponding mRNA in a variegated range of ways and at multiple levels. It can affect the temporal pattern of expression (developmental stage), the level of expression, the cell-type specificity, the capacity to respond to particular cellular or metabolic conditions, the stability of the mRNA, the translation effficiency of the transcript and the structure of the amino-terminus of the protein. This in turn can lead to alterations in protein levels, functions, or subcellular distribution, again all through various mechanisms.

For instance RUNX1 (runt-related transcription factor 1), a key regulator of early hematopoiesis and a frequent target of chromosomal translocations in acute leukemias, is regulated by two widely spaced alternative promoters, separated by more than 160 kbs, that create diversity in distribution and protein-coding potential of the corresponding mRNA isoforms, which are differentially expressed in various cell types, different developmental stages (Levanon and Groner 2004) and during embryonic hematopoiesis (Pozner et al 2007). The RUNX1 mRNA transcribed from the proximal promoter has a long 5`UTR, which contains a functional internal ribosome entry site that mediates a cap-independent translation (Pozner et al 2000). In that case the alternative promoters are used without modifying the protein coding potential of the gene.

Similar examples are represented by OTX2 (orthodenticle homeobox 2) implicated in

medulloblastoma brain tumors and transcribed by three alternative promoters (Curtois et al 2003), and SHOX (short stature homeobox), a cell type–specific transcription factor involved in cell cycle and growth regulation. They use two alterative promoters producing two distinct 5' UTRs (one is longer and highly structured than the other 50 UTR) with identical proteins (Blashke et al 2003), regulated by a combination of transcriptional and translational control mechanisms.

Alternative promoters can also generate mRNA isoforms that encode distinct proteins, sometimes having opposing biological activities, if the translation start site of the corresponding mRNA isoform exists within the first exon.

In most cases, activities of the protein isoforms differ because the promoters are separated by one or two exons that encode an important functional domain.

For example, homer homolog genes Homer1 and Homer2 encode constitutively expressed long protein isoforms (Homer1b/c) and immediate-early response short isoforms (Homer1a) transcribed by alternative promoters (Tapper and Kuner 2006). Homers are synaptic scaffolding proteins, involved in neuronal signaling. Homer1b/c contain a coiled-coil (CC) domain at their C-terminus that enables the cell membrane localization and clustering of glutamate receptors and proteins involved in their intracellular signaling cascades. By contrast, Homer1a lacks the CC-domain and cannot multimerize and thus functions as a natural dominant negative of Homer multimers to influence synaptic plasticità (Tappe and Kuner 2006). Homer1b/c are highly expressed throughout the spinal cord, whereas Homer1a is expressed at very low levels in naive rats (Tappe et al 2006). Homer1a operates in a negative feedback loop to regulate the excitability of the pain pathway in an activity-dependent manner and protects against chronic inflammatory pain.

These isoforms, driven by alternative promoters, differentially regulate cocaine-induced neuroplasticity (Szumlinsky et al 2006).

Another example is LEF1, which encodes lymphoid enhancer factor proteins that mediate the transcriptional regulation of Wnt/b-catenin target genes, is transcribed by two alternative promoters P1 and P2. The most 5` promoter P1 produces a full-length LEF1 protein (LEF1), which recruits β-

catenin to *Wnt* target genes, whereas the intronic promoter P2 drives the shorter LEF1 (ΔNLEF1), which cannot interact with β-catenin and instead suppresses *Wnt* regulation of target genes (Arce et al 2006). Another interesting example is represented by the entire p53 gene family of transcription factors, whose members, p53, p63 and p73 all display to be alternatively spliced and transcribed from alternative promoters (Murray-Zmijewski and Bourdon 2006). It has been shown that collectively all the members of this family have a dual gene structure, essentially determined by the use of two distinct alternative promoters, one upstream of the first exon and the other lying in an intron of the gene, which direct the production of multiple different splicing isoforms with different associated functions.

The discovery of an internal promoter within the p53 family was first made with p63 (Yang et al 1998). The human and mouse p63 genes express at least three alternatively spliced C-terminal isoforms (α, β, γ), and can be transcribed from an alternative promoter located in the intron 3. The transactivating isoforms (TAp63) are generated by the activity of the promoter upstream of exon 1 while the alternative promoter in intron 3 leads to the expression of amino-terminally truncated isoforms (ΔNp63) lacking the transactivation domain. Altogether, the p63 gene expresses at least six mRNA variants which encode for six different p63 protein isoforms (TAp63α, TAp63β, TAp63γ, ΔNp63a, ΔNp63β, and ΔNp63γ). Importantly the TAp63 isoforms are able to bind the same DNA elements (p53RE) as the ΔNp63 isoforms acting as dominant negative forms and inhibiting the function of the proteins containing the transactivation domain. At the same time the ΔNp63 isoforms are still able to activate specific gene targets not induced by the TA isoforms (Wu et al 2003). The TAp63 are expressed very early in the corse of differentiation in the progenitor cells and later in the development a promoter switch occur to change the expression pattern in favour of the ΔNp63 isoforms (McKeon 2004).

In the case of the p73 gene many more isoforms exist, and even more diversity is generated by using different ATG translation initiation start codons further expanding the protein funtional diversification (Moll and Slade 2004). So far, 14 different p73 isoforms have been described.

The p73 gene expresses at least seven alternatively spliced C-terminal isoforms (α, β, γ, δ, ϵ, ζ and Z) (reviewed in Moll and Slade 2004 and Melino et al 2003) and at least four alternatively spliced N-terminal isoforms initiated at different ATG (Stiewe et al 2002). Like p63, the p73 gene can be transcribed from an alternative promoter located in the intron 3. The transactivating isoforms are generated by the activity of the promoter upstream of exon 1 while the alternative promoter in intron 3 leads to the expression of amino-terminally truncated isoforms (ΔNp73) lacking the transactivation domain. Altogether, the p73 gene expresses at least 35 mRNA variants, which can encode theoretically 29 different p73 protein isoforms.

The ΔN isoforms bind DNA through p53RE and can exert dominant-negative effects over p53, p73 and p63 activities either competing for DNA binding sites or by direct protein interaction, showing antiapoptotic characteristics. The p73 variants have been shown in mouse and cell culture studies to have distinct roles. During the normal 'sculpting' of the developing mouse neuronal system, ΔNp73 isoforms are needed to counteract p53- mediated neuronal death.

Moreover, an overexpression of the ΔNp73 isoforms, which inhibits apoptosis, is detected in several tumors including neuroblastoma, lung and ovarian carcinomas.

A similar heterogeneity in alternative promoter usage coupled to alternative splicing and heavy rearrangement in protein domain composition has been seen at p53 locus (Murray-Zmijewski and Bourdon 2006), which is mutated in about 50% of all cancers.

The p53 gene transcription can be initiated from two distinct sites upstream of exon 1 and from an internal promoter located in intron 4. The alternative promoter leads to the expression of an amino-terminally truncated p53 protein initiated at codon 133 (Δ133p53).

The intron 9 can be alternatively spliced to produce three isoforms p53, p53β and p53γ, where the p53β and p53γ isoforms lack the oligomerisation domain. Therefore, the human p53 gene can encode at least nine different p53 protein isoforms, which are named accordingly to p63/p73 nomenclature p53, p53β, p53γ, Δ133p53, Δ133p53β and Δ133p53γ due to alternative splicing of the

intron 9 and usage of the alternative promoter in intron 4, and also Δ40p53, Δ40p53β, Δ40p53γ due to alternative splicing of the intron 9 and alternative initiation of translation or alternative splicing of the intron 2.

p53 isoforms can have distinct biochemical activities. p53β binds preferentially the p53-responsive promoters p21 and Bax rather than Mdm2, while p53 binds preferentially to Mdm2 and p21 rather than Bax promoters (Bourdon et al 2005).

Besides this complex view, one important aspect in the expression regulation by alternative promoters is how the different promoters are chosen. Many factors can contribute to promoter choice in different contexts, among them the diverse core-promoter structure at alternative initiation sites, a variable concentration of *cis*-regulatory elements in the upstream promoter region and local epigenetic mechanisms, such as DNA methylation, histone modifications and chromatin remodeling.

There are few reports showing that such mechanisms influence the choice of which promoter is used. For example, Archey et al. 1999 reported that the lack of methylation within the proximal region of an alternative promoter for TGFB3 (transforming growth factor, b3) correlates with its activity in breast cancer cells. Similarly, a causal link between alternative promoter activity and the silencing of the respective isoforms through histone deacetylation and DNA methylation was demonstrated for the Ras association domain family 1 gene (RASSF1) (Dammann et al 2000) and the p66Shc longevity gene (SHC1) (Ventura et al 2002). Alternative promoters also regulate the genes encoding DNA methyltransferase (DNMT) enzymes, which mediate DNA methylation at various genomic regions in tissue- and developmental stage– specific manner (Schaefer et al 2007). Currently, there are four known families of DNMT enzymes encoded by DNMT1, DNMT3A, DNMT3B and DNMT3L. Most Dnmt genes contain sex-specific germline alternative promoters that are activated at specific stages of gametogenesis. The alternative promoters give rise to germ cell–specific transcripts that lead to the production of shorter forms of the proteins or to untranslated mRNAs that can have regulatory properties (reviewed in Schaefer et al 2007).

Although the functional consequences of Dnmt proteins isoforms are largely unknown, studies suggest that the alternative splice forms modify different chromatin domains and are expressed in a tissue- and developmental-specific manner (Chen et al 2002).

Therefore, cell type–specific expression of Dnmt isoforms and their localization to different chromatin domains might be one mechanism for tissue-specific methylation of alternative promoters of mammalian genes.

Another key component in the epigenetic machinery is the occurrence of histone modifications, which play a crucial role in remodeling of the chromatin state in and around the promoter region of a gene (Guenther et al 2007). A recent genome-wide study that mapped the chromatin state in pluripotent and lineage-committed cells showed that the promoter state reflected the lineage commitment of mouse embryonic stem cells, and alternative promoters have multiple and distinctive chromatin states (Mikkelsen et al 2007), suggesting that an active state at one of the promoters is sufficient to drive the expression of the corresponding transcript.

MYC expression is controlled by two closely spaced but independently regulated alternative promoters. The downstream promoter P2 is the predominant promoter in normal tissues, making up to 75–90% of MYC mRNA, whereas the upstream promoter P1 produces only about 10–25% of MYC mRNA (Marku et al 1992). Amplification and chromosomal translocation of the MYC locus on chromosome 8 to one of the immunoglobulin loci on chromosomes 14, 2 or 22 are characteristic features of Burkitt's lymphoma cells. On chromosomal translocation sites, the insertion of immunoglobulin enhancer elements causes a shift in promoter usage from P2 to P1, resulting in 50–90% of MYC mRNA isoform derived by the P1 promoter (Marku et al 1992).

Disruption of numerous developmentally regulated genes is also implicated in several neuropsychiatric disorders, including Parkinson's disease, schizophrenia, bipolar disorder and autism. There is a growing body of evidence suggesting that most of the developmentally regulated genes produce different splice forms, whose expression in the human brain and other tissues is spatially and temporally regulated through differential transcription from alternative promoters. For

example, several genes including those encoding dopamine receptors (DRD2, DRD3 and DRD4) (Anney et al 2002), serotonin receptor 2A (HTR2A) (Parsons et al 2004), catechol-O-methyltransferase (COMT) (Abdomaleky et al 2006), NMDA (N-methyl-D-aspartate) receptor genes (GRIN2A, GRIN2B, GRIN2C), brain-derived neurotrophic factor (BDNF) (Li uet al 2005) and lysinuric protein intolerance gene (SLC7A7) (Puomila et al 2007) encode different isoforms transcribed by alternative promoters. For most of these genes an association have been proved to the etiology of schizophrenia and related disorders through meta-analyses and large multicenter linkage studies (Abdolmaleky et al 2005).

CAGE data and the analysis of mammalian promoters: Bidirectional promoters

Promoters have been believed to occupy distinct, non-overlapping genomic regions separated by non-functional DNA. Early studies identified some of the exceptions to this rule: genes that lie on opposite strands with their TSSs lying in close proximity to each other form so-called bidirectional promoters (Sugimoto et al 1994, Albig et al 1997, Guargaglini et al 1997) . Trinklein et al in 2004. estimated that 1,352 gene pairs in the human genome have TSSs on the opposite strand that are separated by less than 1 kb; the corresponding number in the mouse was estimated to be 1,638 (Engstrom et al 2006).

Genome-wide analyses that take account of the many newly identified non-coding RNAs and associated CAGE tags have revealed that promoter overlap of this kind is even more common (Katayama et al 2005, Carninci et al 2006). In the large majority of bidirectional promoters, the TSS distribution is of the broad type, although each promoter in a bidirectional pair has independent core promoter elements since the TSS distributions in the two directions generally do not overlap (Carninci et al 2006). Accordingly, Trinklein et al. reported that 23% of bidirectional pairs produce sense–antisense pairs that overlap at their 5′ ends, where the TSS region of one gene is upstream of the TSS region of its partner. Interestingly CAGE data analysis suggests that this is an

underestimate.

Hence, many bidirectional promoters might be more appropriately referred to as anti-directional or opposing promoter pairs.

The genomic distribution of the bidirectional promoters and their associated sense-antisense pairs is not uniform: Sun and collaborators presented evidence for a preferential targeting of 3`UTRs by cis-encoded natural antisense transcripts (cis-NATs) (Sun et al 2005).

Although it is well known that both the 5`- and 3`-UTRs of eukaryotic mRNAs may play a critical role in posttranscriptional gene regulation (Pesole et al 1998, Mignone et al 2002), from the study of Sun and collaborators it is suggested that both *cis*- and *trans*-encoded putative antisense RNAs prefer to bind at the 3`-UTRs of the potential target genes, and such a feature is highly conserved and potentially related to antisense regulation. While miRNA regulation may act on 3`-UTRs in the cytoplasm, 3`-UTR targeted sense-antisense (SA) regulation may occur in the nucleus. In that compartment SA interaction could lead to the formation of double-strand RNAs, activation of RNA editing and nuclear retention of transcripts, modification of chromatin, or other yet obscure consequences (DeCerbo and Carmichael 2005). In agreement with this model, Kiyosawa et al. recently observed that not only is antisense expression widespread, but that a large fraction of natural antisense transcripts are both poly(A) negative and restricted to the nucleus (Kiyosawa et al 2005). Recently, the field of antisense transcription has received much more attention as many different studies demonstrated the huge potential for regulation of the gene expression on a global scale. Identification of NATs in whole genomes is possible due to screening of the large collections of sequence data available from multiple organisms. *In silico* methods for detecting NATs suffer from several shortcomings depending on the source of sequence information (Wang, Gaasterland and Chua 2005). Studies that use mRNAs have sequences whose orientations are known, but the amount of mRNA sequence information available is small (Lavorgna et al. 2004). Predicted gene models using algorithms trained to look for genes gives an increased coverage of the genome at the cost of confidence in the identified gene (Wang, Gaasterland and Chua 2005). Another resource is

the extensive EST libraries but these small sequences must first be assigned an orientation before useful information can be extracted from them. Some studies have used special sequence information in the ESTs such as the poly(A) signal-tail and splicing sites to both filter the ESTs and to give them the correct transcriptional orientation (Osato et al 2007). Combinations of the different sequence sources attempts to maximize coverage as well as maintain integrity in the data.

Pairs of NATs are identified when they form overlapping clusters. There is variability in the cut-off values used in different studies but generally ~20 nucleotides of sequence overlap is considered the minimum for transcripts to be considered as overlapping clusters (Osato et al 2007).

Currently there are a variety of web and software resources that can be used to look for antisense pairs. The Natural Antisense Transcript database (NATsdb) is a rich tool for searching for antisense pairs from multiple organisms (http://natsdb.cbi.pku.edu.cn/). Another web resource publicly available is AntiCode (http://www.anticode.org/) with more than 30,000 NATs from 12 species.

Cis-NATs have a variety of orientations and differing lengths of overlap between pairs (Zhang et al 2006). Five possible orientations for cis-NATs have been identified to date (Fahey, Moore and Higgins 2002). The most common orientation is head-to-head, where the 5' ends of both transcripts align together (Lavorgna et al 2004). This orientation would result in the greatest knockdown of gene expression if transcriptional collision is the reason for transcript inhibition. There is however some studies that have suggested that tail-to-tail orientations are the most common NAT pairs (Osato et al 2007). Others such as tail to tail, overlapping, nearby head-to- head, and nearby tail-to-tail are less frequently encountered. Completely overlapping NATs involve the antisense gene being located completely over top of each other. In head-to-head and tail-to-tail orientations transcripts are physically discrete from each other but are located very close to each other. GO terms enrichment analyses suggests that there is an overrepresentation of NAT pairs in genes that have catalytic activity or implicated in metabolism and cell organization and biogenesis (Zhang et al 2006). There may be something about these genes in particular that makes them more prone to this type of regulation.

Molecular mechanisms behind the regulatory role of cis-NATs are not currently well understood (Lavorgna et al 2004). Three models have been proposed to explain the regulatory effects that cis-NATs have on gene expression. The first model attributes that a 6-8 nt long base pairing between the cis-NAT and its complementary transcript result in a knockdown of mRNA expression (Borsani et al 2005). The assumption of this model is that there will be a precise alignment of at least 6 base pairs between the cis-NAT pair to make double stranded RNA which can give rise to siRNA (Osato et al 2007).

Epigenetic modifications like DNA methylation and post-translational modification of core histones form the basis of the second model (Tufarelli 2006, Osato et al 2007). Although it is not yet clearly understood, it is thought that the reverse transcript guides methylation complexes and/or histone-modifying complexes to the promoter regions of the sense transcript and cause an inhibition of expression from the gene (Tufarelli 2006). Currently it is not known what attributes of cis-NATs are crucial for the epigenetic model of regulation but recently siRNA targeted to gene promoters have been shown than can direct epigenetic modifications that result in transcriptional gene silencing (TGS) in human cells (Han, Kim and Morris 2007).

Further support for the epigenetic modification model comes from the observation that for some non imprinted genes, alterations in the methylation pattern at the promoter regions are demostrated to be directly linked to the synthesis of antisense overlapping transcripts.

Antisense-RNA transcription through the CpG island of a non-imprinted and autosomal tissue specific gene can lead to its silencing and methylation (Tufarelli et al 2003). This phenomenon was observed in a patient with an inherited form of α thalassaemia, whose decreased expression of the alpha globin gene (HBA) is due to changes in chromatin structure such as the hypermethylation of the CpG island (Barbour et al. 2000). In this patient, a small interstitial deletion resulted in the juxtaposition of the intact erythroid specific alpha globin gene to a truncated, widely expressed gene LUC7L, transcribed in the opposite orientation to that of the alpha globin gene (Tufarelli et al. 2001). This results in a 'gene within gene' configuration of the HBA gene on the deleted allele. In

this patient and in a transgenic mouse model, methylation of the alpha globin CpG island is dependent upon the presence of AS-RNA transcripts (HBA-AS) transcribed in *cis* of the methylated gene from the promoter of the truncated LUC7L gene (Tufarelli et al. 2003). The *cis* nature of this AS-RNA mediated DNA methylation event is supported by the observation that the normal allele of the patient is not methylated (Barbour et al. 2000).

It is thought that in malignant cancer cells with activated transposable elements, *cis*-antisense transcription can create a large amount of "transcriptional noise".

It is likely that aberrant antisense RNA transcripts resulting from this transcriptional noise may cause stochastic methylation of CpG islands associated with oncogenes and tumor suppressor genes (Zhang et al 2006). This inhibition would further progress the malignancy of the cells since they lose key regulator genes. By looking at upregulated antisense transcripts in tumor cells, researchers can find more candidate tumor suppressor genes. Also, aberrant *cis*-NATs have been studied as potentially implicated in neurological diseases such as Parkinson`s disease (Liu et al 2005).

The final proposed model that has gained favour due to recent experimental evidence is the transcriptional collision model. During the process of transcription of *cis*-NATs, the transcriptional complexes assemble in the promoter regions of the gene. Rna polymerases will then begin transcribing the gene at the transcription initiation site laying down nucleotides in a 5' to 3' direction (Carmichael 2003). In the areas of overlap between the *cis*-NATs the RNA polymerases will collide and stop at the crash site (Osato et al 2007). In this case, transcription is inhibited because RNA polymerases prematurely stop and their incomplete transcripts get degraded (Røsok and Sioud 2005).

CAGE data and the analysis of mammalian promoters: Mobile promoters

A growing number of experimental data is accumulating showing that mobile elements may drive and modulate the transcriptional activity of the coding part of the genome. Approximately 45% of the human genome is composed of transposable element (TE) derived sequences, compared to

about 1% occupied from the protein-coding sequences, and this is probably an underestimate since many TE sequences would have evolved beyond recognition. TEs are similarly distributed also in other genomes as in human. Their incredible evolutionary success can be explained, at least in part, by their ability to out-replicate the host genome in which they reside, so that TEs can spread within and among genomes, at the expenses of a selective cost to their hosts. This is at the basis of the "selfish DNA" concept which focused on the parasitic nature of those mobile elements, emphasizing the possible deleterious effects on the host (Orgel and Crick 1980). More recently, attention shifted on the potential contributions of TEs to the evolution and transcriptional activity of the host genome. TE sequences seem to contribute to creative and positive evolutionary changes for the host, such as emergine of new alternative promoters, antisense transcription controlling the overlapping sense transcript. There are many cases of TE-induced changes in host gene regulation (Britten RJ 1996, Nekrutenko and Li 2001). A bioinformatics analysis of more than 2000 experimentally determined human promoters have shown that about 24% of these promoters contain some TE-derived sequences (Jordan et al 2003). Most of these TE-derived sequence are present within 3`UTR and 5`UTR of annotated mRNAs, much less along the CDS region.

Two main classes of TE can be distinguished: retrotransposon derived elements (SINEs, LINEs and LTR) and non-coding tandem repeats (satellite DNAs and TTAGGG arrays). SINEs are mainly clustered in gene-rich regions and found in promoters. LINEs are concentrated in gene-poor regions, and more depleted from promoters. Satellite DNAs are mainly located in heterochromatin, implicated in gene silencing. TTAGGG arrays are involved in telomere capping and are also found at centromeres. Both TTAGGG arrays and satellite DNAs can provide binding sites for transcription factors and protect CpG island promoters from repressive chromatin modifications, thus controlling also the boundary among active and inactive sites of transcription in the nucleus (Tomilin 2008). Retrotransposons can act also during development to drive the expression of different isoforms of protein coding genes. In mouse oocytes, LTR class III retrotransposons make an unexpectedly high contribution to the maternal mRNA content, which persist in the embryo.

Many of these mobile elements can act as alternative promoters or first exons for a set of host genes, in a developmentally regulated manner (Peaston et al 2004).

Moreover, other TE-derived RNA species can directly modulate the transcriptional activity of RNA polymerase II holoenzyme, as in the case of SINE B2 which is able to bind RNA polymerase II in response to a heat shock stress and repress transcription of a subset of genes. Noticeably B2 represents the most abundant class of SINEs in mouse (Allen et al 2004, Espinoza et al 2008).

The multitasking genome – a new view for gene expression regulation and transcriptome structure

In molecular biology there are two intriguing paradoxes: the inconsistent relationship between organisms complexity and (1) cellular DNA content and (2) the number of protein-coding genes, referred to as the C-value and G-value paradoxes, respectively. The C-value paradox may be largely explained by varying ploidy. The G-value paradox is more problematic, as the extent of protein coding sequence remains relatively static over a wide range of developmental complexity and evolutionary distances. Recent analyses of sequenced genomes show that the relative amount of non-coding sequences, including introns, increases consistently with complexity (Taft et al. 2007).

Eukaryotes and prokaryotes have evolved substantially different types of genomic architectures: with eukaryotes exhibiting a highly interconnected multitasking organization as opposed to the highly compact proteome-centric bacterial architecture.

Bacterial genome sizes vary from 0.16 Mb as in the case of the obligate endosymbiont *Carsonella ruddii* containing only about 182 protein-coding sequences, and which heavily rely on the host functions, to free-living bacteria with autonomous metabolic and catabolic capacity, such as *Burkholderia xenovorans* (9.7 Mb and 8,602 protein coding genes) (Binnewies et al. 2006).

Prokaryotes contain relatively low amount of non-protein-coding sequences, which mainly consist of 5' and 3' regulatory elements that control transcriptional and translational rates of different cluster of functionally related genes known as operons. In bacteria the transcription and the

translation occur at the same time and in the same compartment and introns are not present within protein-coding genes. Few genes for structural RNAs, like tRNAs, contain some autocatalytic ancient introns (group II introns) which are excised by a different mechanism and without a trans-acting protein complex (Gottesman 2005).

Higher eukaryotes have a much more complex genomic organization, a mosaic gene structure with a dual output, mRNA (protein coding) sequences and introns, which are released from the pre-mRNA by spliceosome-mediated posttranscriptional processing. Additionally, many other transcripts (likely more than half) do not encode proteins at all, but appear to be developmentally regulated and having a genetic function.

Introns have been enormously successful in the course of evolution of eukaryotes, as they comprise up to 95% of the protein-coding genes in mammals.

Interestingly, the distribution of introns in complex organisms is non-random. Genes composed of large amounts of intronic sequence are significantly overrepresented amongst genes that are highly expressed in the nervous system, as well as among genes downregulated in embryonic stem cells. It has been suggested that the informational paradox in complex organisms may be explained by the expansion of *cis*-acting regulatory elements and genes specifying *trans*-acting non-protein-coding RNAs (Taft et al. 2007). Genome sequencing project have revealed that the core proteome sizes of *Caenorhabditis elegans* and *Drosophila melanogaster* are similar and each is only about the double of yeast and some bacteria, despite they appear to have more than twice the complexity of microorganisms, leading to suggest that "the evolution of additional complex traits is essentially an organizational one, a matter of novel interactions that derive from the temporal and spatial segregation of fairly similar components" (Chervitz et al. 1998; Rubin et al. 2000).

Even the human genome has been estimated to contain only about 30,000 protein-coding genes, 99% of which are shared in common with mouse (International Human Sequencing Consortium 2001; Venter et al. 2001). The increase in the complexity of higher eukaryotes is at least in part explained by the production of many different protein isoforms from one single gene by alternative

splicing (Croft et al. 2000). However the most striking feature in the evolution of these organisms was the progressive accumulation of large amount of complex non-protein-coding sequences, which carry up to 98% of the information, as result from the demonstration that most of the genome is actively transcribed (Cheng et al. 2005; Carninci et al. 2005; ENCODE Project Consortium 2007). Moreover, less than 1% of the sequence differences between individual humans occurs in the protein coding sequences (Venter et al. 2001), which suggest that the majority of phenotypic variation between individuals and species results from differences in the control architecture rather than in the proteins themeselves. On the contrary, in bacteria phenotypic variation is primarily achieved by varying the proteome; for instance different strains of *Escherichia coli* have been found to differ by over 20% in their gene complement (Hayashi et al. 2001).

The concept that phenotypic variation and functional complexity in higher eukaryotes arise from differential use of a set of core components started to be represented and summarized by various models, in part inspired to electrical circuits design and mainly focused on the modules rather than on the interconnecting control architecture of the system. Some examples are represented by such concepts as "synexpression groups" (Niehrs and Pollet 1999), "syntagms" of interacting genes (Huang 1998), reuse of modules in signalling pathways (Pawson 1995; T. Hunter 2000), or increased rates of evolution by varying connections between modular network components (Hartwell et al. 1999; Holland 1999).

In particular, network models, ranging in size from simple regulated circuits (Almeida, Fernandez and Infantosi 1998; Mendoza and Alvarez-Buylla 1998; Yuh, Bolouri and Davidson 1998) to entire genomes (Thieffry et al. 1998) have proved that feedback-subnetworks can exhibit computational behaviors like "learned behavior" (Bhalla and Iyengar 1999), that switching networks and transcriptional control networks can exhibit dynamical stability (Wolf and Eeckman 1998; Smolen Baxter and Byrne 2000), and that feedback cinrcuits can implement oscillators governing cell cycles and circadian clocks (Dano, Sorensen and Hynne 1999; Shearman et al. 2000). Usually stochastic noise, time delays allowing feedback, molecular memory and oscillations can be

incorporated in such circuit models generating probabilistic phenotypic variation and amplification of signals (McAdams and Arkin 1997; Smolen, Baxter and Byrne 1999; Hasty 2000).

The major limitation into these models comes from the fact that they are unsuited to describe a real global cellular connectivity and dynamics, because they cannot be scaled up to a large network size, since linear increase in the number of interconnected circuit nodes corresponds to a quadratic increase in the number of interconnecting molecules. This implies an exponential increase in the model size, which severely constrain numerical simulation using the current computing technologies. To bypass this problem, a series of alternative solutions have been sought, by treating subnetworks as active integrated logic components interconnected into a larger network, or considering the system as composed of hierarchically organized control subsystems (McAdams and Shapiro 1995; van der Gugten and Westerhoff 1997).

In an interesting theory formulated by John Mattick and Michael Gagen, the biological systems could have been solved this problem in a different way: adopting a multitasking architecture (Mattick and Gagen 2001). Multitasking is generally referred to as the apparent simultaneous performance of two or more tasks at the same time.

Multitasking is employed in every modern computer in which control codes (program instructions) of n bits set the central processing circuit to process 2^n different operations. In complex organisms, multitasked organization via n controls (single molecules) can, in theory, achieve an exponential (2^n) multitasking of subnetwork dynamical outputs and allow a wide range of programmed responses to be obtained from limited numbers of subnetworks and consequently limited genetic coding information. The obvious advantage and exponential benefit deriving from using a controlled multitasking and the small linear cost of control molecules makes it likely that evolution would have been explored this option.

What multitasking does it mean in practice, when applied to biological organisms? The meaning is obvious for the many different cellular activities which are subject to a fine tuning in time and

space. However, it is not so obvious what are the control molecules that makes it possible for such a system to work.

Why bacterial genomes didn't acquire larger genomes and how higher eukaryotes could develop such a complex organization without exponentially expanding their set of protein coding genes?

Clearly, one of the most striking feature of higher eukaryotes compared to bacteria is the large expansion of the intergenic regions and the appearance of large portions of non coding tracts which separate the protein coding blocks in the pre-mRNAs.

The multitasking hypothesis suggests that it is far easier to modify and expand the number of small control sequences than duplicate and mutate entire subnetworks of genes. Moreover, just turning off controls may reset the program which can be important for survival and reproduction. A control architecture makes it possible to coordinate activity across interacting sets of genes, ensuring a great degree of flexibility and plasticity to the system.

However, a multitasking structure is only useful to the extent that the controls can convey informations about the dynamical state of the network and its surrounding enviroment. To do this, nodes within the network must generate multiple outputs in the form of endogeneous controls. In a fully integrated network, endogenously sourced controls are likely to be more abundant than externally sourceed controls.

Therefore, each gene subnetwork within a cell must produce many different control molecules in parallel with their primary gene products, dynamically communicating with other subnetworks. They may employ all the available biological processes as transcription, splicing, capping and polyadenylation, transport and compartimentalization, decay and turnover, transcriptional and post-transcriptional silencing, translational control, chromatin re-organization, transpositionand its suppression, antisense transcripts production, RNA interference, co-suppression. Interestingly, most of these processes are RNA-mediated events rather protein-mediated events.

Such a system should be able to explore new expression space at fast evolutionary rates over short evolutionary timescales. This fits with the longer average time per generation in the higher eukaryotes, compared to the very short time typically observed in bacteria.

Such networks are generally stable and scale-free, in which some nodes have a high connectivity and others have low connectivity, similar to what is observed in other systems like social communication and internet (Albert, Jeong and Barabasi 2000).

Bacteria have limited genome sizes and low phenotypic complexity, suggesting that advanced integrated organization is not widely employed in these microorganisms. The absence of a prokaryotic multiplex control system also implies that a system built primarily on proteins has inherent limitations. An highly interconnected control system in which endogenously sourced controls were relying on proteins, even just one additional protein ouput per node, would double the size of the genome. Therefore, the multiple outputs needed to have a true dynamical system would require huge increase in the size of the genome and in the energy cost. It is very likely that eukayotes have been used RNA to mediate such endogenous multiple controls: introns and non coding RNAs are both good candidate to play such a role.

For many years the central dogma "one gene-one protein" has biased our perspective in a proteome-centric vision of the biological processes. Only recently, a series of experimental evidences in years of studies, using different approaches in different systems, started to shade new lights on old topics. Introns size and sequence complexity correlates well with developmental complexity.

In developmentally simple eukaryotes like *Schizosaccharomyces pombe*, *Aspergillus*, *Dictyostelium* introns represent only 10-20% of the primary transcripts and are generally small, with an average length of 100 bases and a density of 1 to 3 introns per Kbase of protein-coding sequence.

In the higher plants there are 2 to 4 introns per gene with an average length of 250 bases, spanning for about 50% of the primary transcripts. In animals, a progressive evolutionary increase in the average intron size could be seen, ranging from 500 bases in Drosophila and C. elegans up to 3,400 bases in humans, where there are 6 to 7 introns per gene, representing on average more than 95% of

the primary transcripts (Deutsch and Long 1999; Venter et al. 2001; http://isis.bit.uq.edu.au an extensive introns size database).

Some exeptions are also known to this general trend among higher eukaryotes such as *Fugu rubripes*, which has almost no repetitive DNA in its genome and in which 75% of introns are very small with very few of them accounting for the major part of the unique sequences, and *Arabidopsis thaliana* which also has large part of the introns very small and a compact genome (Brenner et al 1993; Carels and Bernardi 2000).

Small introns are likely to represent vestigial forms, while larger introns characterized by a great sequence complexity may be considered as bearing a functionality. An interesting example is represented by the complex colonial pluricellular alga *Volvox carteri* which have large introns in its genome (Fabry et al. 1993). Since the algae of the order *Volvocales* comprise many closely related species with different levels of complexity, ranging from the simplest unicellular species like *Clamydomonas* to simple colonial forms and even more specialized ones, which resemble a true multicellular organism, this may represent a useful case to test the hypothesis of introns evolution.

While introns generally show less conservation than adjacent protein-coding sequences, some of them are highly conserved over evolutionary distances (Garbe and Pardue 1986; Lloyd and Gunning 1993; John, Smith and Kaiser 1996; Aruscavage and Bass 2000; Sun et al. 2000), and often in large clusters (Jareborg, Birney and Durbin 1999), indicating that they are under functional constraint.

Another point to take into account is that introns are more stable than generally thought.

After splicing, introns, starting from a lariat form, are debranched in a regulated process (Ruskin and Green 1985) but subsequent events are still unknown.

The widespread view of excised introns as highly unstable, and simply discarded and degraded comes from the unjustified a priori assumption that they don`t have any function.

In their theory Mattick and Gagen propose that "it`s very likely that introns are subsequently processed by specific pathways and which generate multiple smaller species which can function

independently as *trans*-acting signals in the network (Mattick 1994) affecting the metabolism of other RNAs and the modulation of the chromatin strucure, among other things".

For instance, it has already been long time ago when it was demonstrated that introns are implicated in the production of small nucleolar RNAs (snoRNAs), a group of hundreds of stable small non coding RNAs localized in the nucleolus and involved in the posttranscriptional modification of ribosomal RNAs and other targets. Most of the intronic snoRNAs are found in introns of different genes coding for ribosomal proteins (L1, L5, L7, L13, S1, S3, S7, S8, S13 amongst the others), ribosome-associated proteins (eIF-4A), nucleolar proteins (nucleolin, laminin, fibrillarin, Nop58, Nop56), the heat shock protein hsc70 and the cell cycle regulated protein RCC1, among others (Prislei et al. 1993; Bachellerie et al. 1995; Maxwell and Fournier 1995; Nicoloso et al. 1996; Filipowicz et al. 1999; Filipowicz et al. 2000). It has also been observed that in some even more specialized cases, some genes have evolved at the point in which their protein coding tracts no longer exist (Tychowski, Shu and Steitz 1996; Bortolin and Kiss 1998; Smith and Steitz 1998; Tanaka et al. 2000; Runte et al. 2001).

On the contrary to what is commonly believed, free excised introns, in both lariat and linear forms, have been found to be abundant (Zeitlin and Efstratiadis 1984) and some are relatively stable (Qian et al. 1992). Furthermore even a relatively insensitive technique like *in situ* hybridization, have shown that, after released from the spliceosome, intronic sequences are detectable in the nucleus, often in a spotted pattern (Xing et al. 1993).

The decay characteristics of these smaller RNA species are likely to provide a molecular memory of pri- and post-activation status, giving a significant efficiency gain over the use of bistable regulated gene networks as memories (Gardner, Cantor and Collins 2000).

Differential decay and diffusion rates as well as compartimentalization may create temporally and spatially complex signal pulses and contribute to a fine control over the cellular activities.

The same considerations, already seen for introns, can also be noticed for other non coding sequences in the genome. The complexity of higher eukaryotes can then be correlated to the extent

of progressively more widespread presence of non coding elements in the genome. This is particularly evident from broad comparative genomics studies. In 1993, the team of Sydney Brenner showed that *Fugu rubripes* genome is only 390 Mb, about one-eighth the size of the human genome, yet it contains a similar repertoire of genes to humans (Brenner et al. 1993). By comparing the human and *Fugu* sequences, common functional elements such as genes and regulatory sequences can be recognized as having been preserved in the two genomes over the course of the 450 million years since the species diverged from their common ancestor. By contrast, non-functional sequences are randomized over this long time period. Over 30,000 *Fugu* genes the great majority have counterparts in human genes*,* and vice versa, with notable exceptions including genes of the immune system, metabolic regulation, and other physiological systems that differ in fish and mammals. In the evolution of vertebrates we can observe a progressive increase in the amount of non coding sequences (Taft, Pheasant and Mattick 2007).

Among large non coding sequences, there are only few scattered islands of functionally relevant and conserved elements, which are good candidates to play a regulatory role (Casillas, Barbadilla and Bergamn 2007; Ponjavic, Ponting and Lunter 2007; Pheasant and Mattick 2007). These islands of highly conserved noncoding sequences (CNSs) have proven to be reliable indicators of functionally constrained sequences

All these evidences together with the recent findings about the complex patterns of transcriptional regulation in mouse as well as in humans have provided unparalleled insight into the complexity of the eukaryotic transcriptome (Carninci et al. 2005; Cheng et al. 2005; Kapranov et al. 2005; Katayama et al. 2005) and suggest that a multitasking model for gene expression regulation has been very likely adopted in the course of evolution.

Then what is a gene?

Language evolves rapidly in contemporary biology.

The term 'gene' was originally coined to define a chromosomal tract that influence a phenotypic trait and was subsequently applied to the genomic elements that produce mRNAs, mostly associated with their protein products (Snyder and Gerstein 2003).

Although such description fits bacterial genes, novel data show that our understanding of mammalian genes and genomes and their annotation are still overly simplistic. Novel data sets show that a large part of the genome is transcribed from multiple positions and in both directions. This will ultimately affect the way we understand and study genes. The multitudes of transcripts that exist have several TSSs, alternative splicing variants and transcripts joining together regions that are traditionally annotated as different genes.

In currently and largely accepted tools used for curated gene annotation, such as RefSeq and LocusLink, priority in selecting the sequence to use as the provisional record is given to the longest sequence that maintains identity through the coding region.

So in this way the representative sequence for a gene is biased from the fact that only the one that maximizes the number of included exons and/or protein domains, is considered valid.

Biologists would be forced to rethink what is a mammalian gene and change the classical textbook definitions, in the light of these multiple evidences. Despite so much transcriptional complexity, it is imperative not only to complete the list of the transcribed parts, their coordinates, controlling elements and timing of expression but also to develop formal ways to treat each one of these.

A first step to do this is to classify the overlapping sequences in TUs (Carninci et al 2005, Okazaki et al 2002), which are further divided into transcriptional frameworks (TKs) if they share TSSs, TTSs or splicing sites. However, this is only a first step in categorizing related transcripts and further logical subdivisions are required that take into account the presence and sequential order of transcriptional landmark elements for subsequent quantitative treatment of transcription.

Conclusion

CAGE largely contributed to change the concept of core promoter itself, showing how on a genome wide scale, many different type of promoters can be distinguished based on the shape distribution of their associated TSS. Two large classes are represented by sharp-type promoters and broad-type promoters. The TATA-box bearing promoters are usually associated to tissue-specific transcripts, whereas CpG island broad promoters are mainly found in housekeeping or ubiquitously expressed genes. Functional promoters have been found not only in proximity of 5`TSS but also in 3`UTR and in internal exons. Many of these in 3`UTR display to have a consensus GGG for the initiator-like motif. Moreover in mammalian genomes there is a widespread occurrence of both alternative promoters and bidirectional promoters, which greatly increase the transcriptional output and variety of mRNAs that can be produced from a locus. Also many mobile genetic elements can drive the transcription of their host protein coding genes.

In the light of all this massive and pervasive, partly overlapping, transcription the current definitions of genes appear to be clearly not suitable anymore, and must be revised in the next future.

Mesencephalic Dopaminergic Cells

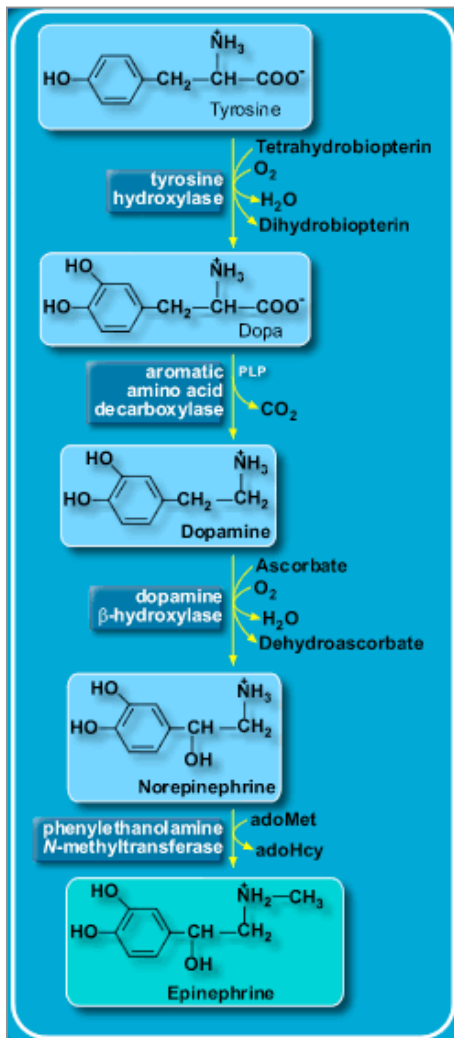## Mesencephalic Dopaminergic Cells: Overview of the system in mammals

In 2007 it has been the 50th annyversary of the discovery of dopamine as an idependent neurotransmitter in the nervous system. DA was first identified as an intermediate in the biosynthesis of noradrenaline and adrenalin from tyrosine (Blaschko 1959). In 1957 Kathleen Montagu at Runwell Hospital near London and Arvid Carlsson with his group in Lund, made the first observations that paved the way for unravelling the role of dopamine as a transmitter in the CNS, demonstrating that DA was present in the brain in about the same concentrations like NA (Montagu KA 1957, Carlsson et al 1958). The marked differences in the regional distribution of DA and NA containing cells led to propose a biological role for DA independent from NA biosynthesis (Von Euler and Lishajko 1957, Bertler and Rosengren 1959).

In 1957–1958, Carlsson and co-workers made the intriguing observation that the akinetic effects of reserpine, a natural alkaloid, could be reversed by an intravenous injection of the dopamine (and noradrenaline) precursor 3,4-dihydroxyphenylalanine (DOPA) and were correlated to a recovery of dopamine, but not noradrenaline, content in the brain, suggesting that depletion of dopamine, rather than noradrenaline or serotonin, was the cause of the akinetic state in reserpine-treated animals. One year later, Carlsson's students Ake Bertler and Evald Rosengren, and Sano and collaborators in Japan, reported that most of the dopamine in the brain was located in the striatum, which contains little noradrenaline, thus providing further support for the hypothesis that this new, putative transmitter might have a central role in the control of motor functions (Bertler and Rosemberg 1959, Sano et al 1959).

This finding, in combination with the Levo-DOPA (L-dopa; 3,4-dihydrox-L-phenylalanine) treatment of PD patients, resulted in the Nobel Prize year 2000 given to Arvid Carlsson, together with Paul Greengard for the investigation of the molecular signaling cascade triggered by

dopamine, and Eric R Kandel who investigated memory formation and learning at the synaptic level.

Dopamine, noradrenaline, and adrenaline are catecholamines, molecules derived from the amino acid tyrosine (Figure 1). The catecholamines (CA) are water-soluble hormones that circulate in the bloodstream, produced either by the adrenal glands or in the postganglionic sympathetic nervous system. In the central nervous system (CNS), dopamine is synthetized by different neuronal populations throughout the brain, where they exert a variety of functions on the basis of their wide connectivity within distinct pathways (reviewed by Prakash and Wurst 2006).

Figure 1. Synthesis of dopamine (DA), noradrenaline (NA), and adrenaline (A). Tyrosine (Tyr.) is converted to dihydroxyphenylalanine (DOPA) by the rate limiting enzyme tyrosine hydroxylase (TH). DOPA is then converted to dopamine by dopa-decarboxylase (AADC). Dopamine-β-hydroxylase (DBH) converts dopamine to noradrenaline, and phenylethanol-amine-Nmethyltran-sferase (PNMT) converts noradrenaline to adrenaline (epinephrine).

The distribution of catecholamines within the brain was mapped in the beginning of the 1960s using the Falck-Hillarp catecholamine formaldehyde histoflouresence method. DA and NA were converted to isoquinoline molecules, which are yellow-green fluorescent (Falck et al 1962). Later the use of immunohystochemical detection of tyrosine-hydroxylase and other specific markers have largely confirmed and expanded the first results (Bjorklund and Dunnett 1984).

Mesencephalic Dopaminergic Cells: Anatomical organization

CA-containing neurons have been found in the brainstem, in olfactory bulb and retina. Dahlström

and Fuxe (1964) described 12 different cell groups, which they designated A1-A12 (Dahlström and Fuxe 1964). Later few new cells were added: the A13 in rostral zona incerta, the A14 cell group in anterior periventricular hypothalamus and the preoptic region (Bjorklund and Nobin 1973) and more recently the A15 cell group in the dorsal preoptic region and in the ventrolateral preoptic and hypothalamic areas (reviewed in Bjorklund and Dunnett 2007a). A16 is assigned to the olfactory bulb group.
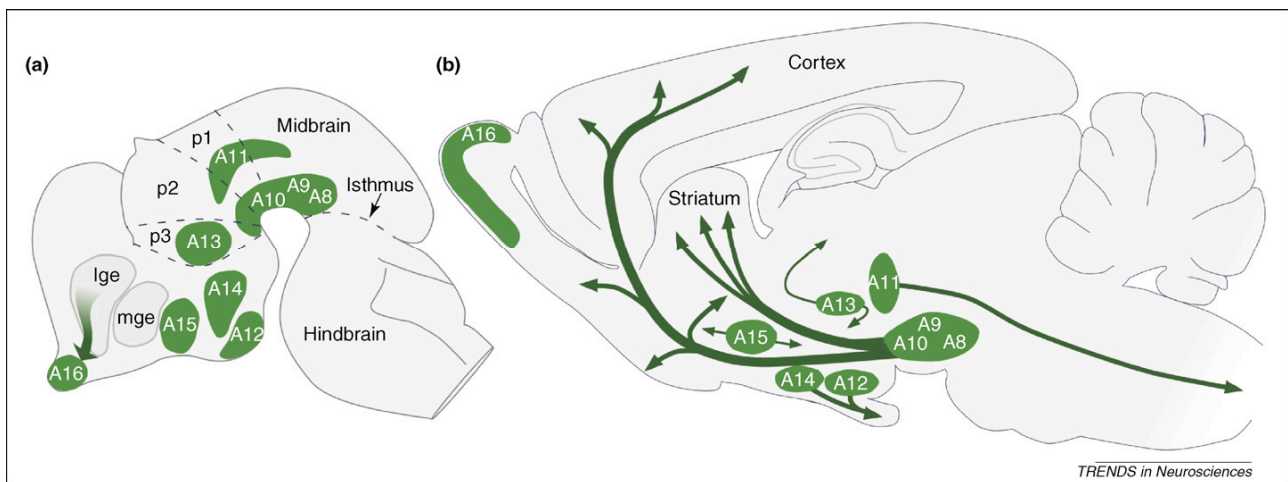

Figure 2. Distribution of DA neuron cell groups in the developing (a) and adult (b) rodent brain. The dopamine neurons in the mammalian brain are localized in nine distinctive cell groups, distributed from the mesencephalon to the olfactory bulb, as illustrated schematically, in a sagittal view, in (a) the developing and (b) the adult rat brain. The numbering of the cell groups, from A8 to A16, was introduced in the classic study of Dahlstro¨m and Fuxe in 1964 [3], and is still valid at present. (Drawing in (a) is modified from Marin et al. [65]; in (b) the principal projections of the DA cell groups are illustrated by arrows). Abbreviations: lge, lateral ganglionic eminence; mge, medial ganglionic eminence; p1–p3, prosomeres 1–3. Taken from Bjorklund and Dunnett 2007

The diencephalic DA cell system (groups A11-A14) comprise neurons which give rise to intradiencephalic projections and to descending diencephalospinal pathway (Figure 2). They are considerably fewer in number compared to the other DA nuclei.

The most prominent assembly of DA-containing neurons are the mesencephalic DA cell system (groups A8, A9 and A10). By means of the tyrosine-hydroxylase immunohystochemistry, the number of mesencephalic DA cells it has been estimated. In rodents, the total number of TH-positive cells in all three cell groups bilaterally is 20,000–30,000 in mice and 40,000–45,000 in rats, with about half of the cells located in the SN. This number increases to between 160,000–320,000 in monkeys and 400,000–600,000 in humans, with >70% of the neurons located in the SN. The

expansion is particularly evident in SN and is accompanied by an extension in the DA innervation territories, in particular in the neocortex, which is much more extensively innervated in primates and human than in rodents (Lewis DA et al 1998, Nelson EL et al 1996). In rodents the cortical DA innervation is limited to frontal, cingulate and enthorinal cortex, whereas in primates it covers almost all the cortical areas (Williams and Goldman-Rakic 1998).

In the brainstem, CA-containing system comprises about 50,000 neurons in the rat, excluding olphactory bulb and retina, out of which 80% are dopaminergic and 20% are noradrenergic. About 5% of the CA neurons are found in medulla oblongata, 15% in pons, 70-75% in mesencephalon and 5-10% in diencephalon (Bjorklund and Hokfelt 1984).

Most of the dopaminergic neurons are highly collaterized: it has been estimated in rat that each nigrostriatal DA neuron have an average total terminal length of about 30 cm, comprising of around 250,000 terminal synaptic boutons (Anden et al 1966). The average content of DA in mesencephalic DA neurons has been estimated to be about 30 pg per neuron (Bjorklund and Lindvall 1984).

Recently, many studies have also focused on the existence of TH immunoreactive cells in various brain regions which apparently do not contain any detectable level of CA (dopamine or noradrenaline). Mainly found in hypothalamus in rodents, and in basal forebrain, striatum and cortical areas in primates, those neurons also lack the decarboxylating enzyme DDC, and the vescicular monoamine transporter VMAT-2 (Ikemoto et al 1996, Ikemoto et al 1999, Weihe et al 2006). Most of these neurons resemble in morphology GABAergic interneurons, and at least in marmosets, they coexpress TH with choline acetyltransferase in magnocellular cholinergic neurons of the nucleus basalis of Meynert (Torres et al 1993).

It is known that during development, the TH gene is expressed in a subset of interneuron precursors that migrate from the lateral and/or medial ganglionic eminences to the cortex, striatum and the olfactory bulb. In rodents, TH protein is transiently expressed in the cortex and striatum only during

the first postnatal month (Berger et al 1985, Satoh and Suzuki 1990). In adult mice and rats, those neurons express the TH mRNA, but no detectable levels of the TH protein, whereas the cells settling in the glomerular layer of the olfactory bulb express both TH protein and synthetize DA, probably as a result of synaptic activation from the olfactory afferents (Baker et al 2003).

Mesencephalic Dopaminergic Cells: Nigrostriatal and Mesolimbic-Mesocortical Pathways

The vast majority of the dopaminergic neurons in the CNS are located in the ventral midbrain and can further be divided into the Retrorubral Field (A8), Substantia Nigra (SN, A9), and Ventral Tegmental Area (VTA, A10) cell groups (Dahlström and Fuxe 1964).

It is often presumed, as a convenient representation, that the mesencephalon contains two major DA neuron subtypes based on their projections: the nigral A9 neurons projecting to the striatum along the nigrostriatal pathway and the A10 neurons of the VTA projecting to limbic and cortical areas along mesolimbic and mesocortical pathways. This has long been recognized as an oversimplification. The SN contains not only neurons projecting to the striatum, but also neurons that innervate cortical and limbic areas; also the DA neurons of the VTA project to the ventral striatum and the ventro-medial part of the caudate-putamen. The A8 cell group consisting of a dorsal and caudal extension of the A9 cell group, contains neurons that project to both striatal, limbic and cortical areas (reviewed by Bentivoglio and Morelli 2005). This intermixing is particularly prominent in primates, much more than in rodents, where the cells of origin of the mesolimbic and mesostriatal pathways are widely distributed throughout the dorsal tier, interspersed among the cells projecting to the striatum (Williams and Goldman-Rakic 1998).

Although this heterogeneity in their established connections with other brain areas, three main functionally distinct pathways can be identified: a mesostriatal pathway, a mesolimbic pathway and a mesocortical pathway.

The "mesostriatal pathway" instead of the "nigrostriatal pathway" could represent a more correct definition to describe the intermixed way in which the anatomical connections are formed by the SN and VTA dopaminergic neurons to reach the two main different striatal compartments.

Nevertheless, in a more restricted sense, when referring to the connections established with the sensorimotor striatum, the classical description of a nigrostriatal pathway formed from neurons of both the ventral and dorsal part of SNc remains valid,.

The nigrostriatal pathway is infact involved in control of motor activity (Ungerstedt 1971).

The loss of the nigrostriatal dopamine projections results in motor disturbances typical of Parkinson's disease, which will be discussed in more detail below.

The mesolimbic pathway consists of dopamine neurons located mainly in the VTA that project to the nucleus accumbens, olfactory tubercle, and amygdala and participate in motivated behavior, reward, and addiction. The VTA togheter with part of SN also projects to the frontal cortex, creating the mesocortical pathway, which is involved in memory and learning (Björklund and Lindvall 1984; Le Moal and Simon 1991, Ungerstedt 1971).

Another useful scheme to classify mesencephalic DA neuron diversity based on connettivity, morphological features and specific molecular markers is the subdivision in dorsal and ventral tiers. In the dorsal tier are located DA neurons of the SN and lateral VTA, with a round or fusiform cell body, innervating ventral striatal, limbic and cortical areas, as well as the matrix compartment of dorsal striatum. These cells express TH together with the calcium-binding protein calbindin, and show to have low level of DAT (Gerfen et al 1987a, Gerfen et al 1987b, German and Liang 1993, Schein et al 1998). The ventral tier is composed of more densely packed angular cells, calbindin-negative but expressing TH together with the G-protein-regulated inwardly rectifying potassium channel 2 (GIRK2) and high level of DAT. Notably, these neurons project almost exclusively to the striatum, in particular to the patch-compartment of the sensorimotor striatal system (Gerfen et al 1987, Prensa and Parent 2001, Thompson et al 2005).

Many of these cells project their dendrites also ventrally into the SN pars reticulata (Prensa and

Parent 2001) and are consequently important for controlling, from within the SN, the output of DA neurotransmittion from the ventral mesencephalon.

Furthermore, the progenitor marker ALDH1 is preferentially expressed in the SN dopaminergic neurons (Chung et al 2005, Haque et al 1997, McCaffery and Drager 1994).

Despite the fact that mdDA neurons in the SNc and VTA have distinct projection targets and functional properties, the population as a whole is often considered to be physiologically and pharmacologically homogeneous. Studies *in vitro* using explants or dissociated neurons prepared from early embrionic tissues, do often distinguish between SN and VTA.

Interestingly, not only the caudate-putamen but also several other basal ganglia structures are innervated by midbrain DA neurons. This includes the external and internal segments of the Globus Pallidus (entopeduncular nucleus in rodents), parts of the ventral pallidum and of the subthalamic nucleus (Lindvall and Bjorklund 1979, Lavoie et al 1989, Gauthier et al 1999).

In the SN itself, DA is known to be released from a plexus of dendritic terminals that is derived from the DA neurons located in the ventral tier of the SNc, and extends throughout large parts of the SNr (Bjorklund and Lindvall 1984). Although the density of their terminals in these nuclei is much less than in caudate or putamen, the ventral tier neurons are strategically placed to regulate some crucial aspects of neurotransmission in the basal ganglia circuitry. Thus, in addition to their potent influence at the level of the caudate nucleus and putamen, the midbrain DA neurons can directly modulate the activity of basal ganglia output neurons at both the pallidal, subthalamic and nigral levels.

In the SN, dendritic DA release has been shown to provide a mechanism by which the nigral DA neurons can regulate not only the activity of the DA neurons themselves, but also the release of GABA within the SNr and the activity of its efferent projections (Robertson HA 1992).

Parent and collaborators (Parent et al 2000, Prensa and Parent 2001) have shown that at least part of the DA innervation in the Globus Pallidus and subthalamic nucleus is derived from collateral branches of DA axons that pass through these nuclei before reaching their final target, the caudate-

putamen. In primates, the internal segment of the Globus Pallidus receives a particularly dense DA innervation, which seems to be derived from a set of neurons located in both SNc and VTA, substantially separate from those innervating the caudateputamen (Lavoie et al 1989, Smith Y et al 1989). This projection is relatively spared in MPTP-treated monkeys and in early cases of Parkinson's disease (Parent et al 1990). Furthermore, Whone and colleagues have observed a compensatory increase in ^{18}F-fluorodopa uptake in the internal segment of the Globus Pallidus in early stage PD patients, which is lost later with the progression of the disease (Whone et al 2003). This would imply the existence of compensatory mechanisms in DA neurotransmission through the Globus Pallidus which can help to maintain a normal motor function when the nigrostriatal pathway begins to be compromised.

Mesencephalic Dopaminergic Cells: The Basal Ganglia circuitry and mDA electrophysiological properties

The basal ganglia system is divided into four distinct brain nuclei; rostrally the striatum (composed of putamen and caudate nucleus), the globus pallidus externa (GPe), the globus pallidus interna (GPi); more caudally the subthalamic nucleus (STN), and the substantia nigra (SN), clasically subdivided in pars compacta (SNc), reticulata (SNr) and lateralis (SNl). The striato-pallido-nigral bundle, which is the core of the basal ganglia, is the anatomical link of the striatum with its primary targets, the pallidum and the substantia nigra. It engages the totality of spiny striatal axons. Estimated numbers are 110 millions in human, 40 milions in chimpanzees and 12 milion in macaques (Percheron et al 1984, Yelnik et al 1987).

These nuclei participate in the control of movements, cognitive, and limbic functions. The motor behavior is controlled in the dorsal striatum, whereas the limbic and cognitive inputs occur in the ventral striatum (Björklund and Lindvall 1984). The dorsal striatum consists of approximately 90-95% γ-aminobutyric acid (GABA)-ergic medium-sized spiny neurons, which receives inputs from

large parts of the motor cortex, thalamus, and the SN (Gerfen and Wilson 1996). Hence, the striatum represents the main input nucleus of the basal ganglia.

Besides the medium-sized spiny neurons, GABAergic medium aspiny interneurons and cholinergic interneurons are also present in the striatum (Gerfen and Wilson, 1996).
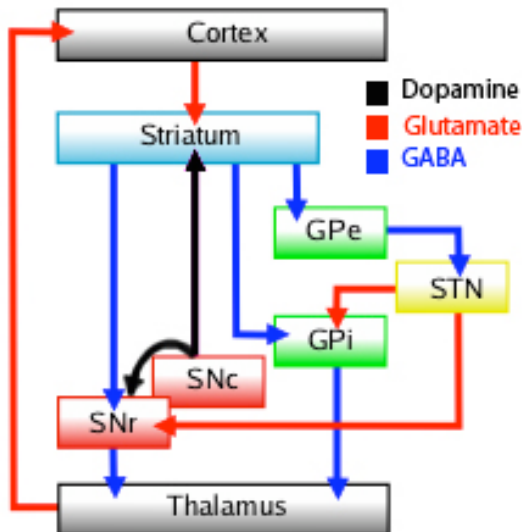

Figure 3. The diagram represents a schematical drawing of the connettivity in the basal ganglia. The dopaminergic imput are colored in black, whereas glutamatergic pathways are in red and GABAergic pathways in blue.

Striatal neurons express dopamine receptors, G-protein coupled receptors that can be divided into two groups: D1-like receptors (receptor D1 and D5) stimulate adenyl cyclase, whereas D2-like receptors (receptor D2, D3, D4) inhibit adenyl cyclase (Dearry et al 1990, Kebabian and Calne 1979, Missale et al 1998). These two distinct functional classes of dopamine receptors are associated to different pathways that characterize the activity of the basal ganglia: the direct stimulatory pathway and the indirect inhibitory pathway. GABAergic medium-sized spiny neurons expressing D1 receptors mediate the direct pathway to the globus pallidus interna, while GABAergic medium-sized spiny neurons expressing D2-receptors mediate the indirect pathway via the globus pallidus externa and the subthalamic nucleus before the signal reaches the globus pallidus interna (Gerfen and Wilson 1996, Graybiel and Ragsdale Jr 1983).

Normal dopaminergic input to the striatum results in an inhibition of the indirect pathway and an activation of the direct pathway. The net effect results in stimulation of the thalamus, which facilitate motor activity. However, in Parkinson's disease, the input of dopamine to the striatum is reduced, which results in a decreased stimulation of the direct pathway and an activated indirect pathway. This results in an increased output from globus pallidus interna and less thalamocortical activity and thus, impaired motor activity (Albin et al 1989, Bjarkam and Sorensen 2004, Blandini et al 2000) (Figure 4).
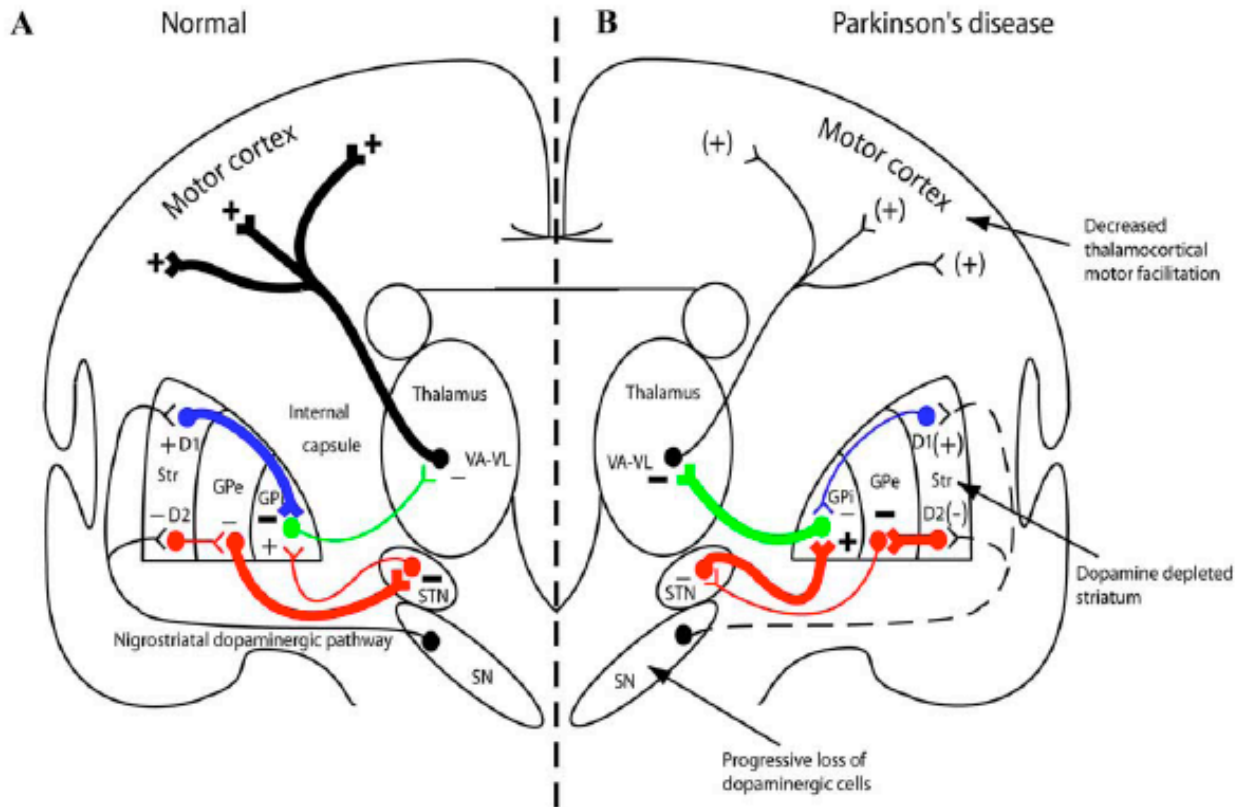
Figure 4. Schematic illustration of the motor basal ganglia system. A) Under normal conditions, dopaminergic input to the striatum facilitate motor activity. B) In Parkinson's disease, the striatum receives less dopaminergic input from the SN, which in the end results in impaired motor activity. The direct pathway is mediated by D1-receptors, and the indirect pathway is mediated by D2-receptors. Striatum (Str), globus pallidus externa (GPe), globus pallidus interna (GPi), subthalamic nucleus (STN), thalamus (Tha). Picture adapted from (Bjarkam and Sorensen, 2004).

Mesencephalic Dopaminergic Cells: Differentiation and Development

The central nervous system derives from the ectoderm germlayer and all the neurons are generated in a region-specific manner from the neuroepithelium by a series of inductive events. Early signalling from organizing centres, such as the midbrain-hindbrain border also known as the isthmus, generates a permissive region that is defined by a specific pattern of gene expression in the mesodiencephalon ventricular zone. This induces mitotic cells in this region to become postmitotic young neurons that are destined to become fully differentiated neurons, acquiring the DA phenotype (Crossley et al 1996, Echelard et al 1993, Hynes et al 1994; reviewed also in Smidt and Burbach 2007, Alavian et al 2008, Sillitoe and Vogel 2008).

First, sonic hedgehog (Shh), together with fibroblast growth factor 8 (Fgf8), defines the specific

region of the isthmus where the mDA neurons will be born. At least other three signalling pathways are involved in these early events and contribute to specify the proper positioning of the isthmus. Wnt signalling pathway (Wnt1 and Wnt5a) is also responsible for the activation of Engrailed transcription factors (En1 and En2) (Danielian and McMahon 1996, Castelo-Branco et al 2003). Transforming growth factor-β (Tgf-β) is essential for the early Shh signalling and subsequent induction of mDA region (Farkas et al 2003). Finally, retinoic acid (RA) is also needed for the correct organization of the midbrain (Avantaggiato et al 1996, Clotman et al 1997).

The isthmus is established very early during development by the mutual repression of two opposing transcription factors, the homolog of drosophila orthodenticle, Otx2, and the gastrulation brain homeobox 2 (Gbx2), thereby defining a sharp border. Wnt1 contributes to the mdDA neuronal phenotype by activating Otx2, which suppresses the expression of Nkx2.2, a transcription factor that is important in inducing the serotonergic neuronal lineage in the hindbrain, and that acts as a suppressor of the dopaminergic lineage (Prakash et al 2006).

The transcription factors engrailed 1 and 2 (En1, En2) have been demonstrated to be essential for the generation and maintenance of mDA neurons (Simon et al 2001, Simon et al 2003, Sgado et al 2006). En1 is widely expressed in mdDA neurons, whereas En2 is expressed in a specific subset of neurons with different ontogeny. En1/2 double knockout mice initially show diminished generation of mdDA neurons, and those that do arise disappear completely later in development (Simon et al 2001). This suggests two different roles for the En genes: one in the generation and/or differentiation and one in the maintenance of mdDA neurons. Expression of α-synuclein by midbrain dopaminergic neurons is diminished in En-1 null mice and absent in engrailed double mutants, although some TH-positive neurons are still present. These findings indicate that En-1 and En-2 regulate the expression of α-synuclein gene in ventral midbrain dopaminergic neurons (Simon et al 2001).

After the expression domain of Otx2 is delineated, and regionalization events have begun, a second set of transcription factors appear: the pair-ruled genes Pax2 and Pax5 (Urbanek et al 1997), the

lim-homeodomain factor Lmx1b (Adams et al 2000) and also the diffusible glycoprotein Wnt1 (Prakash et al 2006). Shortly thereafter, En-1 (at the 1-somite stage) and En-2 (at the 3–5 somite stage) are expressed around the forming midbrain-hindbrain organizer.

Recently, it has been shown that another transcription factor, Foxa2, plays a fundamental role in the specification of mDA precursor cells in the ventral floor plate (Kittappa et al 2007). Foxa2 is a member of the forkhead family of transcription factors, which is known to play a critical role in early development of endoderm and midline structures (Lai et al 1991, Sasaki and Hogan 1994). The expression of Foxa2 takes place when the transcription factors Lmx1b, Nkx2.2, and Phox2a between embryonic day E9.5 and 11.5, defines three adjacent ventral domains of neural progenitors before the first dopamine neurons are formed (Kittappa et al 2007, Wexler and Geschwind 2008) (Figure 5).

Figure5 taken from Kittappa et al 2007 Supplementary Materials
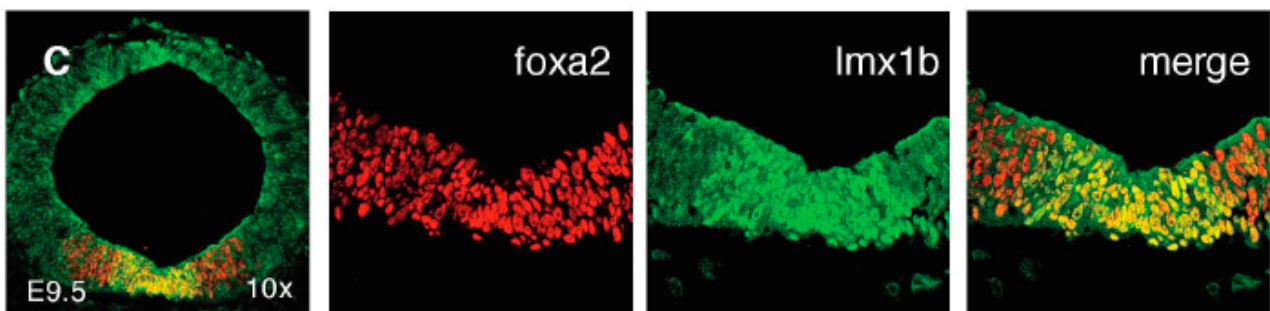
Heterozygous Foxa2[+]/- mutant mice show no gross histological and behavioural defects throughout most of their life, but in late adulthood (18 months aged) about one third of animals begin to develop progressive muscle rigidity with asymmetric posturing and tremors, which are accompanied by an asymmetric selective loss of mDA neurons in SN but not in VTA region.

The first indications of the cellular phenotype of mDA neurons are the appearance of Lmx1a and Msx1 in the still proliferating precursor cells. After this initial round of specification, the precursor cells gradually become postmitotic (in mice from about E10–E14) (Lumsden and Krumlauf 1996) and the first signs of the neurotransmitter phenotype appear, like the expression of the rate-limiting enzyme for dopamine synthesis, tyrosine hydroxylase. At the same stage, several transcription factors begin to be expressed, which are essential for their terminal differentiation and long-term

survival. Deficiencies in Lmx1b, Nurr1, Pitx3, and the Engrailed genes all result in the loss of mDA neurons before birth (Smidt MP et al 2000, Simon et al 2003, Sgado et al 2006) (figure 6). Nurr1 regulates several proteins that are required for dopamine synthesis and regulation, such as tyrosine hydroxylase (TH), vesicular monoamine transporter 2 (VMAT2), dopamine transporter (DAT) and the receptor tyrosine kinase cRET. mdDA neurons are born and are able to develop in the absence of Nurr1. However, they fail to be maintained and have clear defects in transmitter synthesis and release. (Zetterstrom et al 1997, Saucedo-Cardenas et al 1998, Hermanson et al 2003, Smits et al 2003). The Pitx3 gene is expressed in all mDA neurons but in the Aphakia Pitx3-/- mice (which have a 652 bp deletion in the promoter region of Pitx3) the DA neurons in the SNc are lost (Smidt et al 1997, Hwang et al 2003, Nunes et al 2003, Smidt et al 2004), suggesting that Pitx3 could be involved in terminal differentiation or early maintenance of these subset of neurons. Induction of
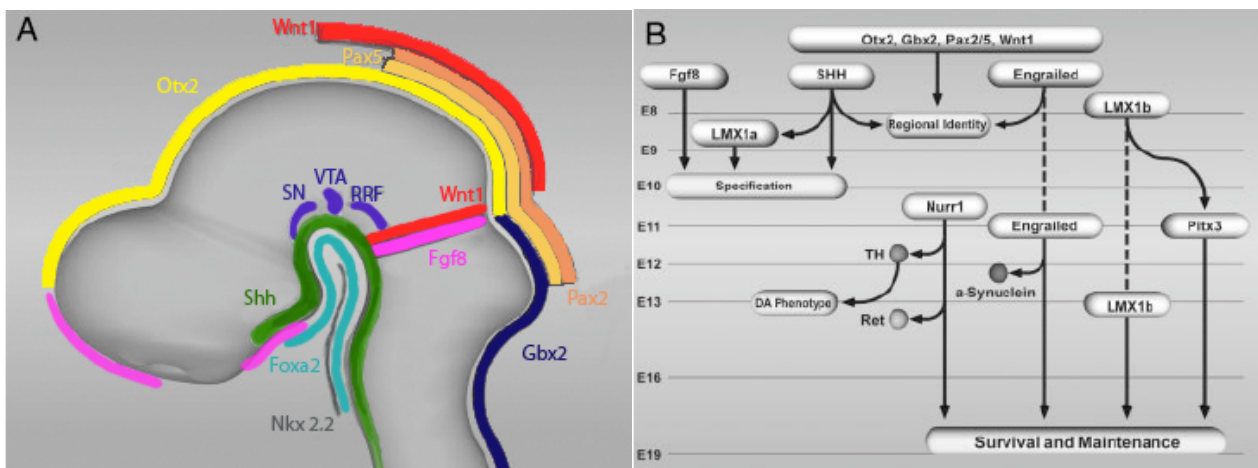

Figure 6. Early and late development, regional identity and induction of mesDA neurons. (**A**) Interaction of sonic hedgehog (Shh), secreted by the floor plate, and the fibroblast growth factor 8 (Fgf8), released from the mid-hindbrain organizer (MHO), induces mesDA neurons in the ventral midbrain. These neurons will later give rise to three distinct nuclei, retrorubral field (RRF), ventral tegmental area (VTA) and substantia nigra pars compacta (SNpc). The location of mesDA neurons is dependent upon the MHO, which is defined by the expression of two opposing factors, *Otx2* and *Gbx2*, and also by *Pax2*, *Pax5*, and *Wnt1*, which participate in the regional specification of midbrain tissue. (**B**) Genes involved in the late development of mesDA neurons. After the early midbrain development and induction of mesDA neurons, the development of postmitotic neurons is controlled by at least three distinct pathways under the influence of *Nurr1*, the *Engrailed* genes, *Lmx1b*, and *Pitx3*. These factors start to be expressed in mesDA neurons around the same time (E10.5-13) and are important for their neurotransmitter phenotype, their survival and maintenance throughout the lifetime of the mouse. Expression of several vital genes is controlled by these transcription factors. In the mice lacking Engrailed genes, the expression of _-synuclein, is significantly reduced. The expression of *tyrosine hydroxylase* (*TH*), the rate-limiting enzyme in dopamine synthesis, and of the *GDNF* receptor component, *Ret*, which is required for long-term survival of nigral DA neurons,[47] is regulated by *Nurr1*. The *Engrailed* genes and *Lmx1b* are expressed in the region before induction of mesDA neurons (dotted lines); yet, their expression becomes relatively specific to postmitotic mesDA neurons during later embryogenesis (continuous lines).

Pitx3 in mouse embrionic stem cells was also shown to lead to an up-regulation of genes highly expressed in nigral DA cells, like aldehyde dehydrogenase 2 (Ahd2), suggesting a role of retinoids during development, in part controlled by Pitx3 (Chung S et al 2005). Pitx3 also seems to control

transcription of many DA-phenotype related genes, like TH and DAT (Lebel et al 2001, Maxwell et al 2005) and a coordinated action with Nurr1 seems to occur in many cases (Martinat et al 2006). Moreover mir133b controls in a negative feedback loop the action of Pitx3 (Kim et al 2007).

At the same time when the dopaminergic precursor cells become postmitotic, they start to migrate along radial glia towards their final location (Shults et al 1990). In the rat, the development of VM dopaminergic neurons has been thought to occur between embryonic day (E) 11-15, with the peak at E14 (Altman and Bayer 1981).

However, recent findings have demonstrated that the vast majority of the substantia nigra (SN) dopaminergic neurons are born already at E12 (Gates et al 2006). When the SN neurons have reached their positions in the midbrain, they form axons projecting towards the lateral ganglionic eminence (LGE), which develops into mature striatum. In the rat, TH-positive nerve fibers are found in the LGE already at E14, however, the dopamine innervation is not completed until weeks later (Smidt and Burbach 2007). Comparing the rat with humans, the ventral migration of TH-positive cells in humans starts at approximately gestation week 7 and the first TH-positive nerve fibers are observed in the developing striatum at gestation week 9, although it takes years before the striatum is fully innervated (Freeman et al 1991).

Different sets of molecules have been demonstrated, by means of different transgenic and knockout mice, to act as axon guidance cues in the migration of the young mDA neurons: Netrin1 (Ntn1) together with its receptor, deleted in colorectal cancer (Dcc) (Fazeli et al 1997, Flores et al 2005). Reelin has been implicated in the functional organization of the mDA system (Ballmaier et al 2002); the proteoglycan phosphacan 6B4, proposed to act together with adhesion molecule L1CAM (Ohyama et al 1998); the expression of neural cell adhesion molecule (NCAM) and polysialic acid has been reported for migrating mDA neurons (Shults et al 1992).

Furthermore, an extensive axonal pruning has been demonstrated to occur during the late stages of development to refine and consolidate the topographical organization of the established mesostriatal and mesolimbic connections (Hu et al 2004). Pruning is often used to selectively remove exuberant

neuronal branches and connections in the immature nervous system to ensure the formation of functional circuitry (Luo and O'Leary 2005). In contrast to the adult, axon collaterals originating from embryonic VTA and SNc neurons (E15 and E17) do not display a preference for the dorsal or ventral striatum (Figure 7). The molecular signals that regulate the selective pruning of mDA axon collaterals remain to be identified.
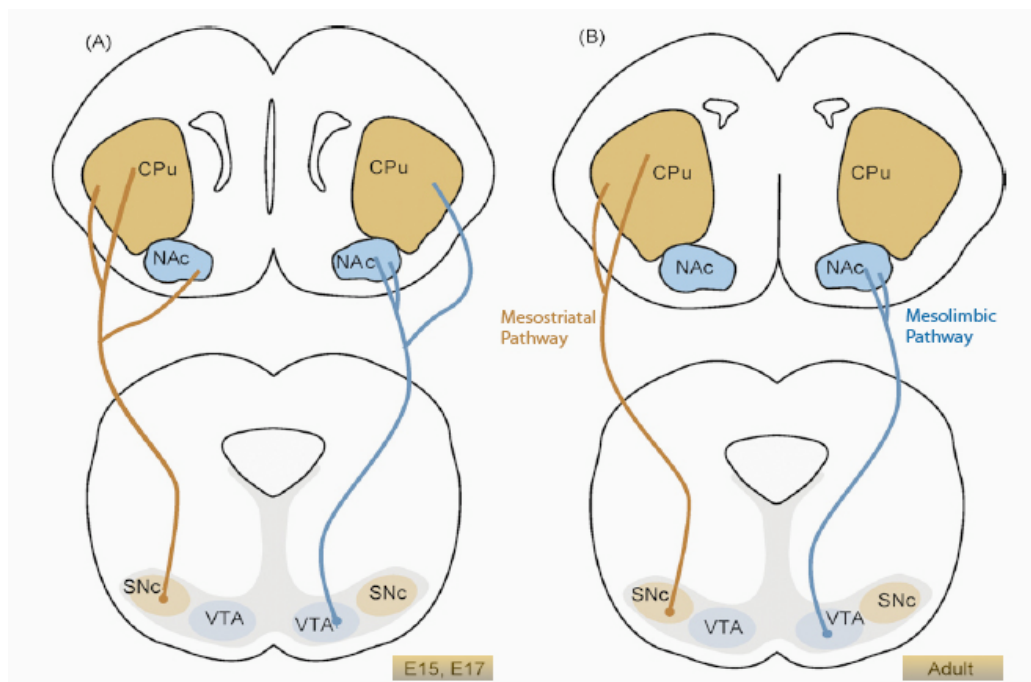
Figure 7. At stage E15- E17, mDA axon collaterals from mDA SN or VTA neurons do not display any preference for the dorsal or ventral striatum; the topographical specificity of mesostriatal and mesolimbic projections observed in adult is achieved during late embryonic and early postnatal development by the selective elimination of VTA and SNc axon collaterals. Picture modified from Hu et al. 2004.

Finally, once neurons have migrated to their final destination, they elaborate axons and dendrites along predetermined routes in the developing embryo to establish highly specific connections with their targets. This requires the proper growth and guidance of extending axons and dendrites as well as their subsequent branching and pruning, with the formation of functional synapses. Many other classes of molecules have been found to participate in these processes –among them ephrins and their receptors Ephs (EphB1, EphA4, EphA5, EphB3, EphA7; Liebl et al 2003, Yue et al 1999, Chung et al 2005, Grimm et al 2004, Willson et al 2006), semaphorins, plexins and neuropilins (Sema3A, Sema7A, plexinC1; Kawano et al 2003, Chung et al 2005, Pasterkamp et al 2007); and the repelling proteins Slits and their Robo receptors (Slit1. Slit2, Slit3, Robo1 and Robo2; reviewed by Dickson and Gilestro 2006).

Mesencephalic Dopaminergic Cells: DAT, VMAT2 and COMT are crucial components of their function

Dopamine Transporter (DAT): function, expression and regulation

It was about four decades ago when the concept of neurotransmitter reuptake was introduced (Hetting and Axelrod 1961) to explain how sympathetic nerve terminals are replenished by NA.

It has been proposed that the reuptake process was an important mechanism for inactivating neurotransmitters. Soon afterwards the discovery of NA reuptake, similar but distinct uptake mechanisms were shown to exist for dopamine and 5-hydroxytryptamine (5-HT, serotonin) (Iversen 1971). This led to the discovery that certain antidepressant and psychostimulants block the reuptake of monoamines in the brain. During the 1990s advances in molecular cloning techiniques led to the isolation of genes coding for monoamine transporters.

The gene encoding the DAT was first cloned in 1991 by several groups (Kilty et al 1991, Giros et al 1991, Shimada et al 1991, Usdin et al 1991). This transporter belongs to the Na+/Cl⁻-dependent family of neurotransmitter transporters, which also includes transporters for the related biogenic amines norepinephrine and serotonin (NET and SERT respectively), as well as for the inhibitory neurotransmitters GABA and glycine (Masson et al 1999). All of these transporters have a common topological arrangement consisting of 12 transmembrane domains (TMs), a large glycosylated loop between TMs 3 and 4, and intracellular amino and carboxy terminal domains (figure 8).

DAT can be considered a specific marker for dopaminergic neurons because it is expressed exclusively in neurons that synthesize DA as a neurotransmitter. There is increasing evidence suggesting that DAT exists as an oligomeric complex in cells. The data are consistent with a model in which oligomerization is necessary for the efficient exit of DAT from the endoplasmic reticulum (Sorkina et al 2003, Torres et al 2003a). Several TMs have been postulated to be involved in the assembly of this oligomeric complex (Hastrup et al 2001, 2003; Torres et al 2003a). To carry out its role, DAT must be targeted to specialized domains near sites of presynaptic DA release.
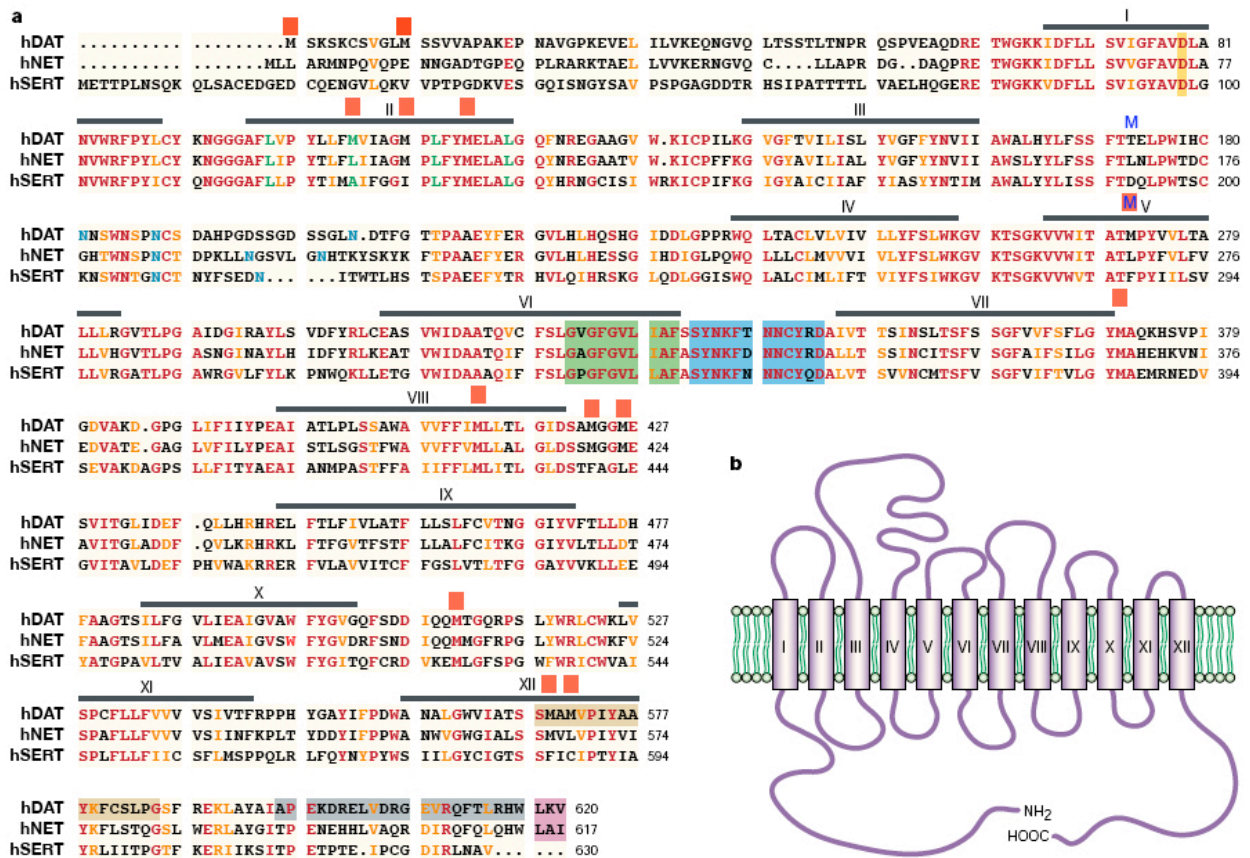
```
a
                                                                                                        I
hDAT  ..........  .........M SKSKCSVGLM SSVVAPAKEP NAVGPKEVEL ILVKEQNGVQ LTSSTLTNPR QSPVEAQDRE TWGKKIDFLL SVIGFAVDLA 81
hNET  ..........  ......MLL  ARMNPQVQPE NNGADTGPEQ PLRARKTAEL LVVKERNGVQ C....LLAPR DG..DAQPRE TWGKKIDFLL SVVGFAVDLA 77
hSERT METTPLNSQK QLSACEDGED CQENGVLQKV VPTPGDKVES GQISNGYSAV PSPGAGDDTR HSIPATTTTL VAELHQGERE TWGKKVDFLL SVIGYAVDLG 100
                                    II                                  III                                    M
hDAT  NVWRFPYLCY KNGGGAFLVP YLLFMVIAGM PLFYMELALG QFNREGAAGV W.KICPILKG VGFTVILISL YVGFFYNVII AWALHYLFSS FTTELPWIHC 180
hNET  NVWRFPYLCY KNGGGAFLIP YTLFLIIAGM PLFYMELALG QYNREGAATV W.KICPFFKG VGYAVILIAL YVGFFYNVII AWSLYYLFSS FTLNLPWTDC 176
hSERT NVWRFPYICY KNGGGAFLLP YTIMAIFGGI PLFYMELALG QYHRNGCISI WRKICPIFKG IGYAICIIAF YIASYYNTIM AWALYYLISS FTDQLPWTSC 200
                                                         IV                                              M V
hDAT  NNSWNSPNCS DAHPGDSSGD SSGLN.DTFG TTPAAEYFER GVLHLHQSHG IDDLGPPRWQ LTACLVLVIV LLYFSLWKGV KTSGKVVWIT ATMPYVVLTA 279
hNET  GHTWNSPNCT DPKLLNGSVL GNHTKYSKYK FTPAAEFYER GVLHLHESSG IHDIGLPQWQ LLLCLMVVVI VLYFSLWKGV KTSGKVVWIT ATLPYFVLFV 276
hSERT KNSWNTGNCT NYFSEDN... ...ITWTLHS TSPAEEFYTR HVLQIHRSKG LQDLGGISWQ LALCIMLIFT VIYFSIWKGV KTSGKVVWVT ATFPYIILSV 294
                                                                 VI                                        VII
hDAT  LLLRGVTLPG AIDGIRAYLS VDFYRLCEAS VWIDAATQVC FSLGVGFGVL IAFSSYNKFT NNCYRDAIVT TSINSLTSFS SGFVVFSFLG YMAQKHSVPI 379
hNET  LLVHGVTLPG ASNGINAYLH IDFYRLKEAT VWIDAATQIF FSLGAGFGVL IAFASYNKFD NNCYRDALLT SSINCITSFV SGFAIFSILG YMAHEHKVNI 376
hSERT LLVRGATLPG AWRGVLFYLK PNWQKLLETG VWIDAAAQIF FSLGPGFGVL LAFASYNKFN NNCYQDALVT SVVNCMTSFV SGFVIFTVLG YMAEMRNEDV 394
                            VIII
hDAT  GDVAKD.GPG LIFIIYPEAI ATLPLSSAWA VVFFIMLLTL GIDSAMGGME 427
hNET  EDVATE.GAG LVFILYPEAI STLSGSTFWA VVFFVMLLAL GLDSSMGGME 424
hSERT SEVAKDAGPS LLFITYAEAI ANMPASTFFA IIFFLMLITL GLDSTFAGLE 444
                           IX
hDAT  SVITGLIDEF .QLLHRHREL FTLFIVLATF LLSLFCVTNG GIYVFTLLDH 477
hNET  AVITGLADDF .QVLKRHRKL FTFGVTFSTF LLALFCITKG GIYVLTLLDT 474
hSERT GVITAVLDEF PHVWAKRRER FVLAVVITCF FGSLVTLTFG GAYVVKLLEE 494
                           X
hDAT  FAAGTSILFG VLIEAIGVAW FYGVGQFSDD IQQMTGQRPS LYWRLCWKLV 527
hNET  FAAGTSILFA VLMEAIGVSW FYGVDRFSND IQQMMGFRPG LYWRLCWKFV 524
hSERT YATGPAVLTV ALIEAVAVSW FYGITQFCRD VKEMLGFSPG WFWRICWVAI 544
                           XI                            XII
hDAT  SPCFLLFVVV VSIVTFRPPH YGAYIFPDWA NALGWVIATS SMAMVPIYAA 577
hNET  SPAFLLFVVV VSIINFKPLT YDDYIFPPWA NWVGWGIALS SMVLVPIYVI 574
hSERT SPLFLLFIIC SFLMSPPQLR LFQYNYPYWS IILGYCIGTS SFICIPTYIA 594

hDAT  YKFCSLPGSF REKLAYAIAP EKDREL.VDRG EVRQFT.LRHW LKV 620
hNET  YKFLSTQGSL WERLAYGITP ENEHHLVAQR DIRQFQLQHW LAI 617
hSERT YRLIITPGTF KERIIKSITP ETPTE.IPCG DIRLNAV... ... 630
```

Figure 8. *Amino acid sequence and topology of monoamine transporter proteins.* **a** | Amino acid alignment of the human dopamine transporter, noradrenaline transporter and 5-HT transporter. Identical residues are shown in red, whereas similar residues (V, L, and I, K and R, F and Y, or D and E) are shown in orange. Asparagine residues that form part of N-linked glycosylation consensus sequences are shown in blue. A conserved aspartate residue in transmembrane domain 1 that is presumably involved in the interaction with monoamines is shown as a yellow box. A leucine-repeat in transmembrane domain 2, and a glycophorin-like motif in transmembrane domain 6, are shown in green. The intracellular loop between transmembrane domain 6 and 7 contains several residues involved in conformational changes during substrate binding and translocation and is shown in blue. Black bars represent putative transmembrane domains. The colour boxes covering the parts of the intracellular carboxyl termini represent interacting sites with Hic-5 (beige), synuclein (grey), and PICK1 (purple). DAT, dopamine transporter; h, human; NET, noradrenaline transporter; SERT, 5-hydroxytryptamine transporter. The samll red boxes over hDAT represent in frame methionines **b** | Proposed topology of monoamine transporters depicting 12 transmembrane domains connected by intracellular and extracellular loops. Picture modified from (Torres et al 2003). The M in blue represent the position of the internal TSS seen in A10 neurons in mouse.

Immunoelectron microscopy studies have revealed that DAT molecules are not located at sites where neurotransmitter release takes place, but are confined to perisynaptic areas (Nirenberg et al 1996), which implies that specific targeting mechanisms must exist for the proper localization of this transporter. Interestingly, these studies also demonstrated that DAT proteins are associated with intracellular compartments, particularly membranes of tubulo-vesicular structures. These findings supported the notion that DAT molecules are strategically targeted to perisynaptic sites in nerve terminals suggesting that a recycling mechanism might exist to shuttle the transporter proteins between intracellular compartments and the plasma membrane.

At the cell membrane, the function of DAT can be regulated by multiple second messenger systems,

including protein kinase A, protein kinase C (PKC), protein kinase G, tyrosine kinases, phosphatases, calcium and calmodulin-dependent kinases, and arachidonic acid (reviewed in Doolen and Zahniser 2001, Mortensen and Amara 2003, Melikian 2004, Vaughan 2004). The best characterized effect is the down-regulation of the transporter activity by PKC activators, which has been described extensively in several systems, including striatal synaptosomes and heterologous cells. PKC activation down-regulates DAT by decreasing the transporter Vmax with little change in substrate affinity. This modulation occurs largely through redistribution of the transporter protein at the cell surface rather than changes in intrinsic transport activity. In PC12 cells stably transfected with DAT, PKC activation increases the rate of transporter internalization while decreasing the rate of transporter recycling to the plasma membrane (Loder and Melikian 2003). However, in MDCK cells expressing DAT, activation of PKC leads to internalization and subsequent degradation of the transporter (Daniels and Amara 1999). More recent studies suggest that trafficking of DAT can also be regulated by substrates and inhibitors of the transporter. DA and amphetamine induce internalization of DAT in heterologous cells (Saunders et al 2000), whereas cocaine increases cell surface expression of the transporter (Daws et al 2002, Little et al 2002). Together, these results suggest that rapid DAT endocytic trafficking mechanisms are established at neuronal synapses. Indeed, a non-classical endocytic signal has been identified recently in the most proximal region of the intracellular C-terminal of DAT (Holton et al 2005) and multiple residues in the same region of the transporter are responsible for export from the endoplasmic reticulum (Miranda et al 2004).

However, to date very little is known about the regulatory components involved in trafficking of DAT. Several proteins have been reported to physically interact wiht DAT (Torres 2006).

These interactions suggest that the synaptic distribution, targeting, compartmentalization, trafficking and functional properties of DAT can be regulated via interacting proteins. Most of the known DAT-interacting proteins have been identified using the yeast two-hybrid (Y2H) system. In this assay, protein–protein interactions can be detected by cotransforming the bait of interest with a library of clones and growing the transformants in yeast media selective for the genes expressed as

a result of the interaction (Fields and Song 1989). The DAT-interacting proteins identified to date using the Y2H system are: protein interacting with C kinase- 1 (PICK1) (Torres et al 2001), the focal adhesion protein Hic-5 (Carneiro et al 2002), SNAP-25 (synaptosome associated protein 25 kDa) (G. Torres, unpublished observations), α-synuclein (Lee et al 2001), receptor for activated C kinase-1 (RACK1) and syntaxin (Lee et al 2004). Alternative approaches such as co-immunoprecipitation resulted in the identification of the protein phosphatase PP2A (Bauman et al 2000) and PKC-bII (Johnson et al 2005) as DAT interactors.

Biochemical and functional studies have clearly demonstrated that the DAT exists as an oligomeric complex in cells and identified several TMs involved in transporter assembly (Hastrup et al 2001, 2003, Torres et al 2003a). Hastrup et al. (2003) demonstrated that human DAT could be cross-linked as a dimer at the plasma membrane of human embryonic kidney (HEK)293 cells providing evidence for a role of TM4 and TM6 in oligomerization process. Mutant transporter molecules that fail to oligomerize are retained in the endoplasmic reticulum (Torres et al 2003a, Sorkina et al 2003).

A leucine repeat in TM2 appears to be important for the initial steps in assembly of DAT. Mutation of the leucine repeat in TM2 resulted in a transporter protein devoid of N-linked glycosylation, which suggests that DAT assembly precedes glycosylation during trafficking of the transporter to the plasma membrane (Torres et al 2003a). Based on these results we postulated a role for oligomerization in endoplasmic reticulum processing and cell surface trafficking (Figure 9).

Cocaine and amphetamines, two important psychostimulants, interact with dopamine transporter.

Cocaine and other chemically related drugs are nonselective, competitive inhibitors of monoamine transporters (Ritz et al 1987). By contrast, amphetamine-like drugs are substrates for monoamine transporters (Sulzer et al 1995, Jones et al 1998). Once inside the synaptic terminal, these drugs act as weak bases at synaptic vesicles, causing the redistribution of vesicular monoamines into the cytoplasm, and a reversal in the direction of neurotransmitter transport at plasma membrane monoamine transporters (Sulzer et al 1995, Jones et al 1998). As a result, the application of

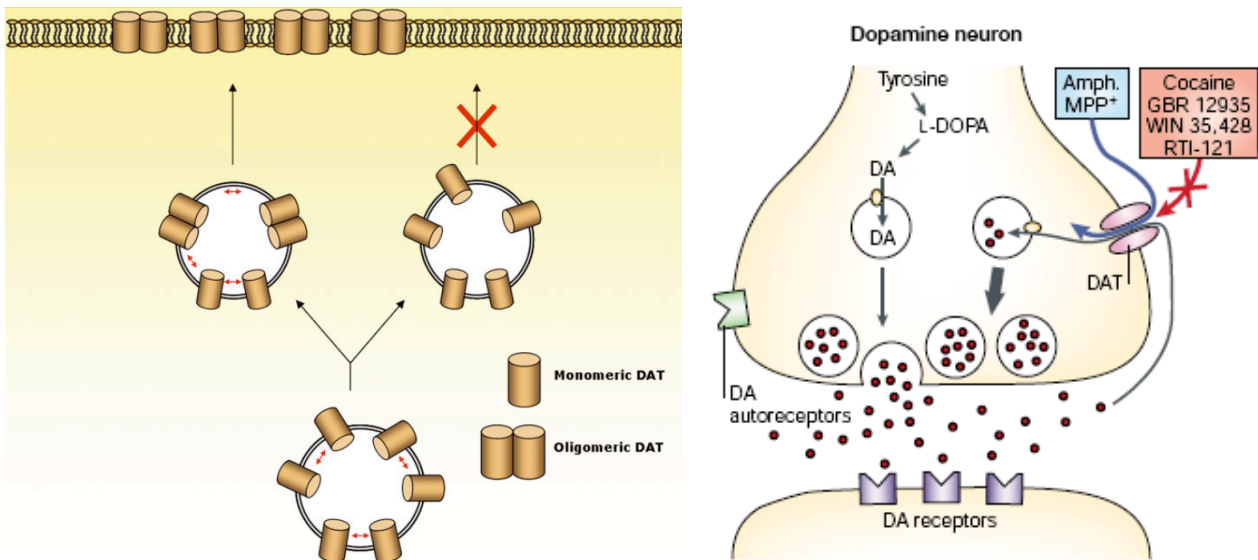amphetamines induces a massive release of monoamines into the extracellular space.

Figure 9. Dopamine transporter is subjected to form oligomeric complexes. The model proposes that DAT assembly into an oligomeric complex is required for trafficking of the transporter complex to the cell surface. Mutant transporters that fail to oligomerize are retained in the ER. The DAT oligomers are mainly localised in presynaptic densities in lateral positions. Metamphetamines stimulate the DAT activity whereas cocaine and synthetic derivatives inhibits the reuptake.

DAT: multiple transcript isoforms

At least for the human SLC6A3 gene, there are accumulating experimental evidences that multiple transcript isoforms are produced from this locus, as summarized in the following scheme (figure 10).
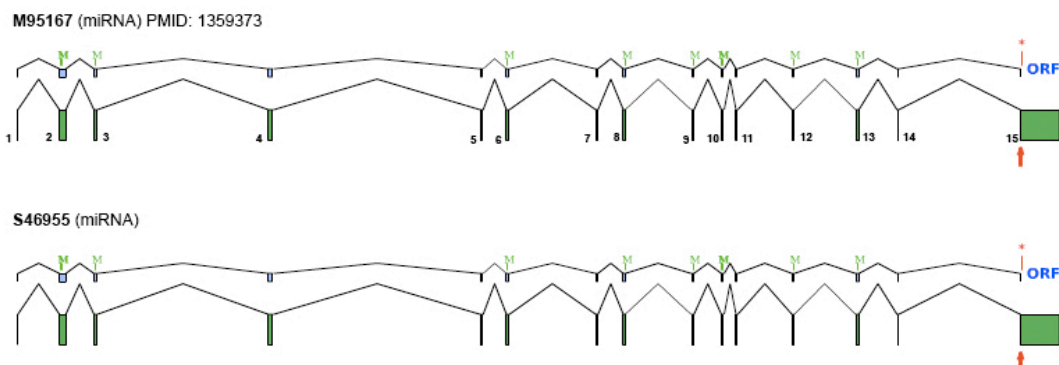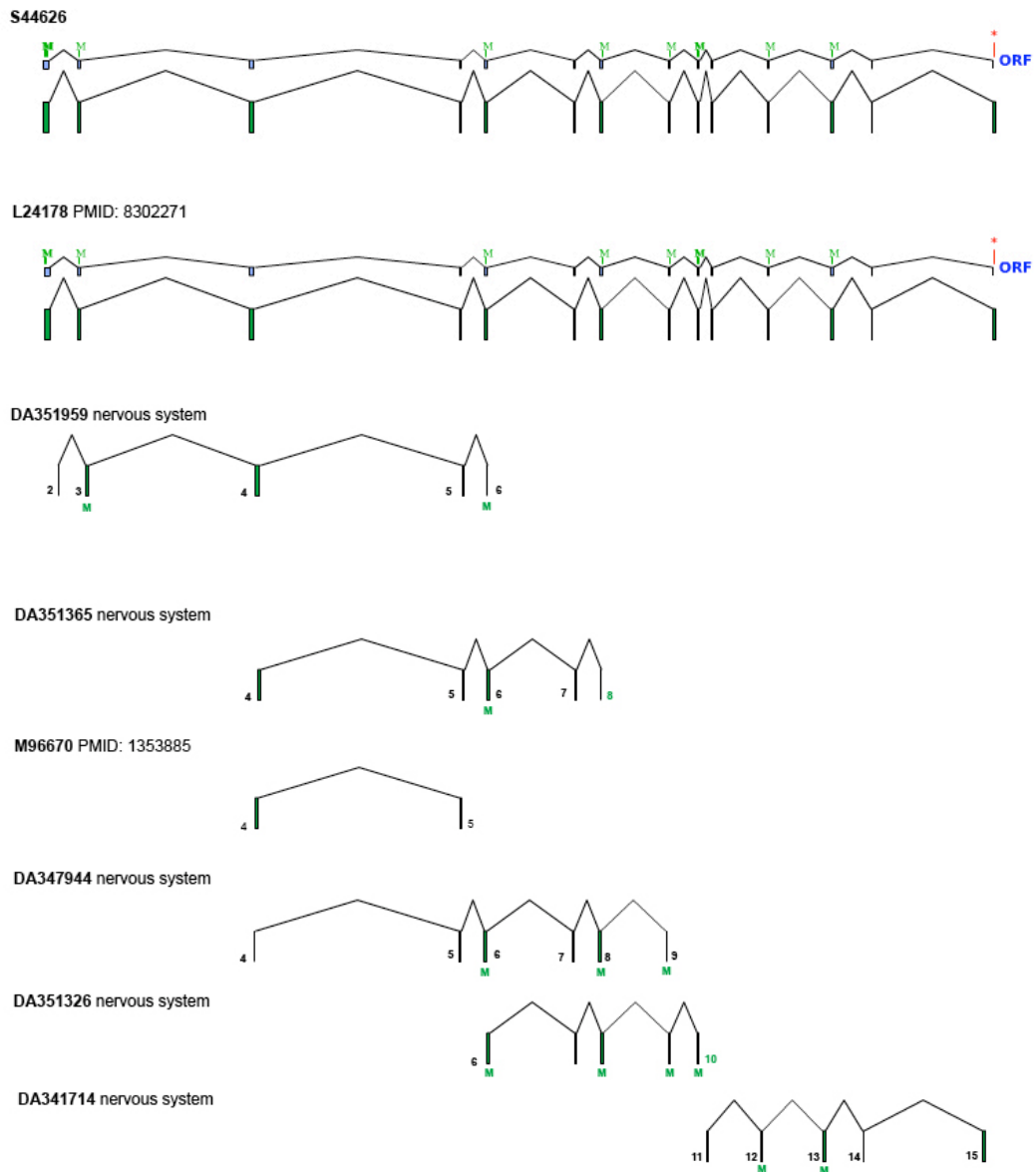
Figure 10. Summary of all the known EST / cDNA clones at the human SLC6A3 locus. Most of them derives from oligo-capped cDNA from human substantia nigra (Kimura et al., Genome Res. 16-1: 55-65, 2006). - M are methionines.

Vmat2: function, expression and regulation

At the presynaptic site, availability and fusion competence of synaptic vesicles as well as the vesicular transmitter content contribute to the strength of postsynaptic responses.

Vesicular monoamine transporters (VMATs) translocate monoamines from the cytosol into the secretory vesicles of monoaminergic neurons, neuroendocrine cells, and platelets (Liu et al 1992).

Transport is driven by an electrochemical proton gradient ($\Delta\mu H^+$) across the vesicular membrane,

which is generated by a vacuolar H^+-ATPase (Kanner and Schuldiner 1987). In mammals two closely related isoforms of the monoamine transporter, termed VMAT1 and VMAT2, respectively, were identified (Kanner and Schuldiner 1987, Erickson et al 1992). The transporter proteins presumably contain 12 transmembrane domains that are located on different vesicle subtypes (Erickson et al 1992, Nirenberg et al 1995). Both VMATs transport serotonin, dopamine, epinephrine, and norepinephrine but differ in their substrate preferences and affinities. In contrast to VMAT2, VMAT1 prefers epinephrine over norepinephrine, and the *Kd* for serotonin uptake is around 1 μM for VMAT1 but below 1 μM for VMAT2. Furthermore, histamine is only transported by VMAT2. The activity of both transporters is irreversibly inhibited by reserpine, whereas tetrabenazine (TBZ) exclusively inhibits VMAT2 (Peter et al 1994, Erickson et al 1996).
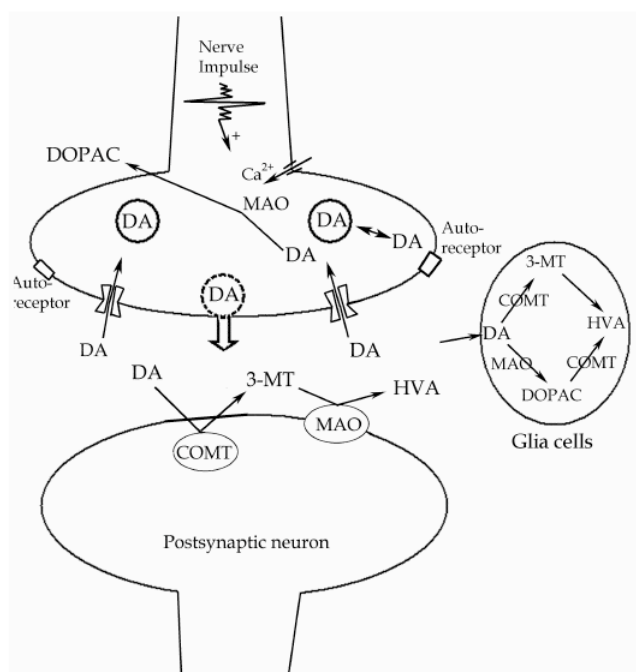
VMAT2 mRNA is localized to monoaminergic neurons and VMAT2 protein is largely localized to synaptic vesicular membranes (Gonzalez et al 1994, Peter et al 1995, Weihe et al 1994).

VMAT2 function is important for proper packaging of monoamines for quantal, calcium-dependent vesicular release (Henry et al 1994, Johnson 1988, Kanner and Schuldiner 1987, Njus et al 1986). In addition, it may be important for sequestering toxins. The active metabolite, 1-methyl-4- phenyl-phenydium (MPP$^+$) of the selective dopamine neuronal toxin *N*-methyl-1,2,3,6-tetrahydropyridine (MPTP) kills dopaminergic neurons or other cells that express the plasma membrane dopamine transporter, and causes a parkinsonian syndrome *in vivo* (Burns et al 1983, Heikkila et al 1984, Pifl et al 1993). Overexpression of VMAT2, however, can suppress the toxicity of MPP$^+$ by sequestering the toxin in vesicles away from presumed sites of cytoplasmic mitochondrial damage (Liu et al 1992, Liu et al 1994). Amphetamines cause cytoplasmic release of monoamines from vesicular stores through disruption of the pH gradient across vesicular membranes (Sulzer and Rayport 1990, Rudnick and Wall 1992). The balance between this release and VMAT2-mediated vesicular reuptake may be important for the non-quantal, calcium-independent extracellular dopamine release that modest amphetamine doses can yield, and for the dopaminergic toxicity that high-dose amphetamines can exert, perhaps through cytoplasmic oxidative stresses with dopamine

oxidation and free radical production (Seiden et al 1993, Zheng et al 2006).

COMT: function, expression and regulation

The catechol-O-methyltransferase (COMT) enzyme inactivates circulating catechol hormones, catechol neurotransmitters and xenobiotic catecholamines by methylating their catechol moieties (Axelrod and Tomchick 1958, Rivett and Roth 1982, Rivet et al 1983, Jeffery and Roth 1984). COMT has a role in the biosynthesis of melanin (Pavel 1993) and several investigators have postulated the involvement of this enzyme in such human mental disorders as depression and schizophrenia (Murphy and Wyatt 1975, Abdolmaleky et al 2008). COMT also inactivates L-Dopa, a catechol-containing drug used as a dopamine precursor in the treatment of Parkinson's disease (Ball et al 1972, Guldberg and Marsden 1975). Due to its harmful effect on the medication of Parkinson's disease, a great deal of interest has been directed towards the development of specific COMT inhibitors (Guldberg and Marsden 1975, Mannisto and Kaakkola 1989) and the determination of the structure of the enzyme itself. Catechol-*O*-methyltransferase (COMT) inactivates dopamine by catalyzing the transfer of a methyl group from *S*-adenosyl-L-methionine to dopamine, generating 3-methoxytramine (3-MT) (figure 11).

Figure 11. Two different uptake processes terminate the synaptic action of released catecholamines in brain: the high-affinity uptake to presynaptic nerve terminals (uptake$_1$, followed by oxidation by monoamine oxidase, MAO) or glial cells uptake (uptake$_2$, followed by *O*-methylation by COMT, and/or oxidation by MAO). For dopaminergic neurons, uptake by the high-affinity dopamine transporter (DAT) is the most effective mechanism, and the contribution of glial COMT remains secondary under normal conditions. Picture modified from Cooper 2003

COMT protein and enzyme activity are widely distributed in mammalian brain (Lundstrom et al

1995, Mannisto and Kaakkola 1999). In the striatum, the synaptic action of dopamine is thought to be largely terminated by neuronal uptake by abundant dopamine transporters (Giros et al 1996). In the prefrontal cortex, however, dopamine transporters are expressed at low levels within synapses and the rate of dopamine uptake is slow (Garris et al 1993, Sesack et al 1998, Lewis et al 2001, Wayment et al 2001). Thus, in the prefrontal cortex, one might speculate that alternative mechanisms, such as degradation by COMT, might be a key process in the regulation of dopamine availability. Consistent with this notion, studies of COMT knockout mice have demonstrated that dopamine levels are increased in the prefrontal cortex of these mice, but not in the striatum, suggesting a more critical role for COMT in cortical versus subcortical sites (Gogos et al 1998).

COMT enzyme occurs in mammals as two distinct forms: in the cytoplasm as a soluble protein (S-COMT), and in association with membranes as a membrane-bound form (MB-COMT). S-COMT activity is the most prevalent in all tissues, while MB-COMT activity represents generally less than *5%* of the total COMT activity (Jeffery and Roth 1984, Guldberg and Marsden 1975, Grossman et al 1985). However, in some human tissues, like the brain, the amount of MB-COMT activity has been reported to be higher (Rivett et al 1983). Two COMTspecific transcripts, 1.3-kb S-mRNA and 1 .5-kb MB-mRNA, are synthesized from two separate promoter regions (Tenhunen et al 1994).

The relative amounts of the COMT mRNAs vary in different human tissues and cell lines, suggesting a tissue-specific regulation of the expression. The two COMT promoters can be regulated by the methylation status: a CpG hypermerthylation specifically inactivates MB-COMT without affecting the S-COMT promoter in endometrial cancer (Sasaki M et al 2003).

Comt: multiple transcript isoforms

Recently, much attention has received the transcriptional regulation at the COMT locus, and some studies have been focusing on the presence of alternative mRNA variants for COMT transcripts.

In one study Tunbridge and colleagues focused attention on potential mRNA variants present in the

brain. They detected 7 COMT variant mRNAs, resulting from both insertions and deletions within the known COMT brain transcript. Several of the variants alter the predicted coding sequence. Three of these variants correspond to sequences within the Aceview database and could be reliably amplified, while the remaining four do not correspond to any expressed sequence tags and were amplified only once (Tunbridge et al 2007). Here you can find a diagram containing all the variants (Figure 12).
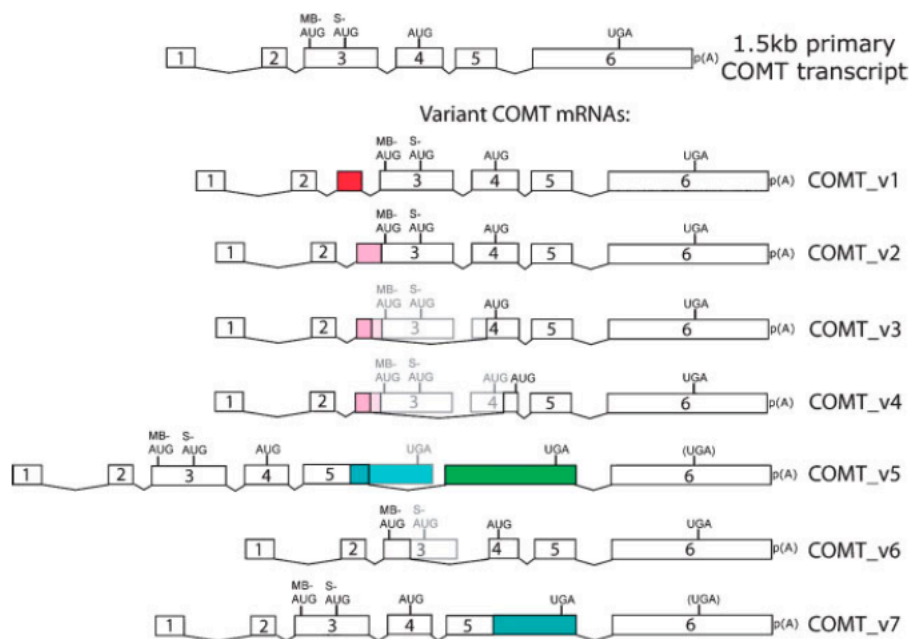

Figure 12. The primary COMT mRNA transcript and variants. The sequence of the primary COMT mRNA is shown, with individual exons numbered and represented by boxes. The variantCOMTmRNAsare shown below and are contrastedwith the primary sequence: additional sequence is indicatedwith color (each color indicates a particular additional exon; note that some pink sequence is shared byCOMTvariants 2–4 and some blue sequence by variants 5 and 7). Faded exons and features indicate deleted sequence. The position of stop codons (UGA), and selected AUG start codons are indicated. All variant mRNAs contain the functional Val158Met polymorphism, which is located near the 30 end of exon 4. Picture taken from (Tunbridge et al 2007)

Mesencephalic Dopaminergic Cells: previous gene expression studies

The catecholaminergic cell system has been largely studied and extensively characterized from the anatomical, developmental and electrophysiological point of view, employing a combination of different approaches ranging from anatomical reconstruction with fluoro-labeled retrogade tracers, immunocytochemistry, reverse transcriptase-polymerase chain reaction of aspirated cytoplasm, patch-clamp whole cell recordings which gained a fairly global picture of the main features that distinguish anatomically distinct group of cells. With the advent of more high-troughput

technologies, like cDNA/oligo array platforms and high-throughput, partially automated, *in situ* hybridization (ISH) systems, more sophisticated tools became available for a more detailed and comprehensive description of cell diversity at the molecular level.

One of the first hightroughput studies of gene expression profiling designed to uncover mesencephalic DA neuron diversity have been conducted in 2004. A microarray analysis has been applied to four different catecholaminergic neuron populations (SN, VTA, LC and A13), in their normal physiological condition, isolated from adult rat brain by laser capture microdissection (Grimm et al 2004). 200 cells from each population were used in 4 independent replicates, and the RNA amplified by two rounds of T7-based linear amplification was hybridized to 14,800 element rat cDNA microarrays. A common pool of amplified RNA from total brain was used as a common reference to evaluate the enrichment of set of genes, already known to be highly and specifically expressed in each CA population: tyrosine hydroxylase (Th) expressed in all CA cells, dopamine beta hydroxylase (Dbh) restricted to NA locus coeruleus cells, GTP cyclohydrolase I (Gtpch1), pterin-4-a-carbinolamine dehydratase (Pcd) and vescicular monoamine transporter 2 (Vmat2). Using a molecular phylogenetic approach, it was shown that SN and VTA neurons differed in only 122 (<1%) of their expressed genes; on the contrary SN and A13 neurons had >5% of significantly differentially expressed transcripts, and the differences were even more when considering SN versus LC (>7%) and LC versus A13 (>10%). These results were somehow expected, considering that SN and VTA are functionally and developmentally more close to each other than with all the other groups.

Despite their similarities, there are some critical differences in gene expression between SN and VTA neurons, both at the level of individual genes and in several potentially relevant cellular pathways. In general, unbiased gene expression profiling has successfully confirmed nearly every previously noted difference in gene expression between the regions (Table 1 and 2), proving that the large datasets produced by profiling experiments can be reliably used for developing testable hypotheses (Grimm et al 2004, Chung et al 2005, Greene et al 2005).

Table 1. Several known differentially expressed genes confirmed using large scale expression profiling

Higher expression in SN	References	Higher expression in VTA	References
D2 dopamine receptor	Hurd et al 1994	Calbindin d28k	Liang et al 1996
mGluR1	Shigemoto et al 1992	BDNF	Seroogy et al 1994
Parvalbumin	Alfahel-Kakunda and Silverman 1997	Neurotrophin 3	Seroogy et al 1994
Glutamate decarboxylase	Hedou et al 2000	Neuromedin K receptor	Massi et al 2000
IGF1	Garcia-Segura et al 1991	alpha-1-beta-adrenergic receptor	Andersson et al 1994
GIRK2 ion channel	Schein et al 1998	Cholecystokinin	Seroogy et al 1994
Aldehyde dehydrogenase 1A7	McCaffery and Drager 1994	GTP cyclohydrolase I	Dassesse et al 1997

Table 2. Recent independently identified genes differentially expressed between SN and VTA neurons by three microarrays stusies (Grimm et al 2004, Chung et al 2005, Greene et al 2005)

Higher expression in SN	References	Higher expression in VTA	References
Sncg Gamma-synuclein	Grimm et al 2004, Green et al 2005, Chung et al 2005	MARCKS Myristolated alanine-rich C-kinase substrate	Grimm et al 2004, Green et al 2005, Chung et al 2005
NMDAR2C N-methyl-D-aspartate receptor subunit 2C	Grimm et al 2004, Green et al 2005, Chung et al 2005	ADCYAP1 Pituitary adenylate cyclase activating polypeptide	Grimm et al 2004, Green et al 2005, Chung et al 2005
		LPL Lipoprotein lipase	Grimm et al 2004, Green et al 2005, Chung et al 2005

modified from Greene JG 2006

An unsupervised clustering analysis revealed relationship among classes of CA neurons, defining at some extent, the molecular signature of each class of CA neurons. First the authors described genes that are expressed in common with all the CA groups: among those pan-CA transcripts, the most represented categories were genes involved in neurotransmitter synthesis and transport, scanvenger of stress-induced cell damage, inflammation related genes, cell-adhesion molecules, different isoforms of the aldehyde dehydrogenase family and modulators of DA-receptor activity.

Gene filtering and multiclass significance analysis revealed 534 genes whose expresssion differed between at least two groups of CA neurons. Several interesting individual genes (122) were differentially expressed between SN and VTA neurons in addition to those previously identified.

Among these factors were transcripts from various categories, including transcriptional regulators (Sox-6, Zfp 288, HTF, and NGFI-A), molecules involved in vesicle trafficking (DOC2B, rab3B, and MARCKS myristoylated alanine-rich C kinase substrate), axon guidance (neuropilin 1, slit-2, and ephrinB3), ion channels [CLIC5 (chloride intracellular channel 5), VR1 (vanilloid receptor 1), and NMDAR2C], transporters (VGLUT2 and CNT2), and G protein-coupled receptors (α-1B-adrenergic receptor and GPRC5C) The two most prominent functional classes encoded factors involved in synaptic plasticity and in cell survival and protection. Strikingly, most of these

transcripts were expressed at a higher level in the VTA neurons.

Table 2 list five genes independently confirmed as differentially expressed in three separate large-scale microarray experiments (Grimm et al 2004, Chung et al 2005, Greene et al 2005). Considering that these experiments were performed on different species, with different techniques, and on different microarray platforms, this is an extremely stringent list.

MARCKS (myristoylated, alanine-rich C-kinase substrate) has recently been found to be an important regulator of dendritic spine plasticity through an activity and PKC-dependent pathway (Calabrese and Halpain, 2005). These functions of MARCKS are important for learning and memory and the plastic synaptic events that underlie them, such as long-term potentiation (LTP) (Matus 2005). Interestingly, MARCKS gene expression is altered in postmortem brains from suicide patients, and its expression is down-regulated by the mood stabilizer lithium (McNamara et al 1999, Wang et al 2001). Since dysfunctional VTA dopamine neurons are thought to be involved in addiction behaviour and mood disorders, higher MARCKS expression in this neuronal population provides a tantalizing clue for further investigation. Pituitary adenylate cyclase activating polypeptide (ADCYAP1 or PACAP) is known to support survival of dopamine neurons, and lipoprotein lipase may protect cells from damage caused by oxidized lipoproteins (Takei et al 1998, Paradis et al 2003, Reglodi et al 2004). Both are intriguing in that VTA dopamine neurons are relatively protected from degeneration in parkinsonism. The neuronal role of γ-synuclein is not clear, but it has been reported to be involved in regulation of the cell cycle, abnormalities of which have been implicated in neurodegeneration (Inaba et al 2005). NMDAR2C may possibly be involved in excitatory neurotoxicity in SN dopamine neurons (Kress and Reynolds 2005). Several single gene products discovered by expression profiling to be higher in the VTA have been successfully tested *in vitro* against dopaminergic neurotoxins, including the aforementioned PACAP, as well as gastrinreleasing peptide (GRP) and calcitonin/calcitonin gene-related peptide alpha (CGRP) (Chung et al 2005). Furthermore, an analysis of the associations to categories of genes revealed that there are concerted differences in gene expression between different

subpopulations of dopamine neurons (Figure 13), and it is likely that these broad categorical and pathway differences account for the distinctiveness of the neuron types (Chung et al 2005, Greene et al 2005). Profiling experiments have confirmed that neuropeptide genes as a category are more highly expressed in VTA dopamine neurons than in those from the SN (Chung et al 2005, Greene et al 2005). The other most prominent categorical difference between SN and VTA dopamine neurons is energy metabolism. Multiple genes related to energy pathways, electron transport and mitochondria are more highly expressed in the SN than the VTA (Chung et al 2005; Greene et al 2005). This is particularly interesting because mitochondrial dysfunction may be critical to the pathogenesis of PD.
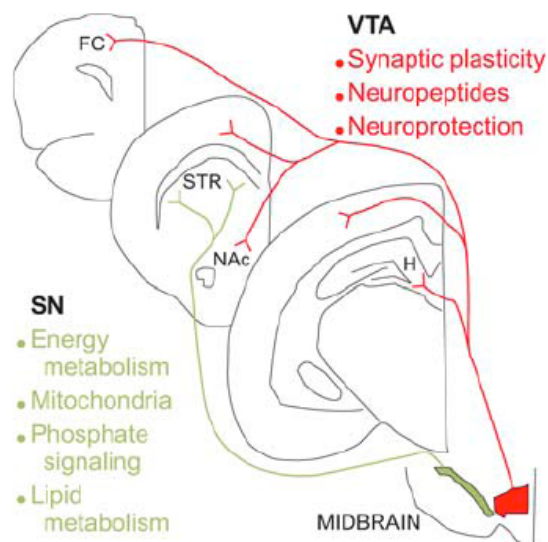
Figure 13   - picture taken from Greene 2006

Mesencephalic Dopaminergic Cells: neuropsychiatric disorders

Even though SN and VTA neurons are located in close proximity, develop in response to the same inductive signals, and have partially overlapping axonal projection domains, they mature to innervate distinct targets, control different brain functions, and display different vulnerabilities to neurodegeneration and other phatologies. While SN neurons are selectively affected by movement disorders associated to neurodegenerative conditions such as Parkinson`s disease, VTA neurons are more linked to the development of mood disorders such as schizophrenia, bipolar disorder,

123

attention-deficit hyperactivity disorder (ADHD), depression and drugs addiction.

Traditionally the VTA was thought to be composed of two main neuronal populations, the DA neurons (or primary neurons) representing the majority of cells, and γ-aminobutyric acid (GABA)-ergic neurons, acting merely as local circuit neurons that make synapses onto neighbour DA cells (Johnson and North 1992, Lacey Mercuri North 1989). Recent studies however have suggested that VTA contains at least three main subpopulations on the basis of their electrophysiological (intrinsic conductances), pharmacological (response to DA, opioids and serotonin) and biochemical (presence of TH) properties (Margolis et al 2003). For this reason, VTA neurons are now distinguished in primary DA-ergic neurons, secondary GABA-ergic neurons, and tertiary neurons whose neutransmitter identity has still to be characterized (Figure 14).
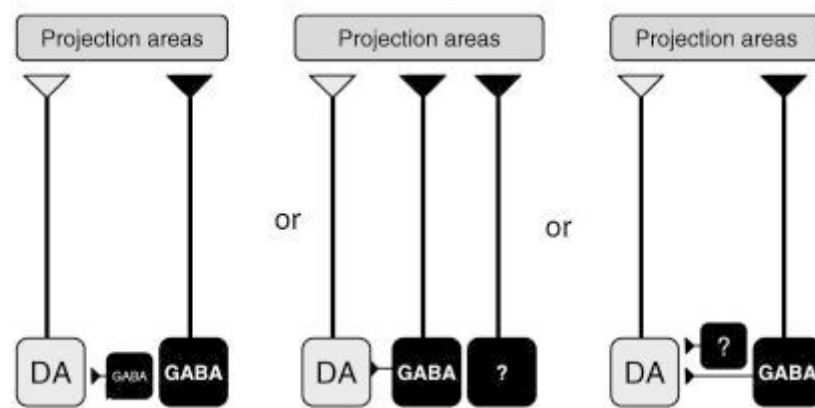

Figure 14. VTA different neuronal subpopulations - picture taken from: Miller Cummings "The Human Frontal lobes"

The activity of VTA DA-ergic neurons is under the physiological control of several neurotransmitter systems from different afferent projections – Figure 15 (Grillner and Mercuri 2002). These include GABA, from local GABA-releasing cells and from GABA-ergic projections from the basal ganglia; glutamate from several brain regions, including PFC, hippocampus, amigdala, and pontine nuclei; ascendine biogenic amine systems, comprising noradrenaline from locus coeruleus, serotonin from the raphe nuclei, acetylcholine from pontine nuclei, and several neuropeptides (Phillipson 1979). Importantly, dopamine itself is released somatodendritically within VTA (Beckstead et al 2004) and the balance of afferent imput to the VTA and the activation

124

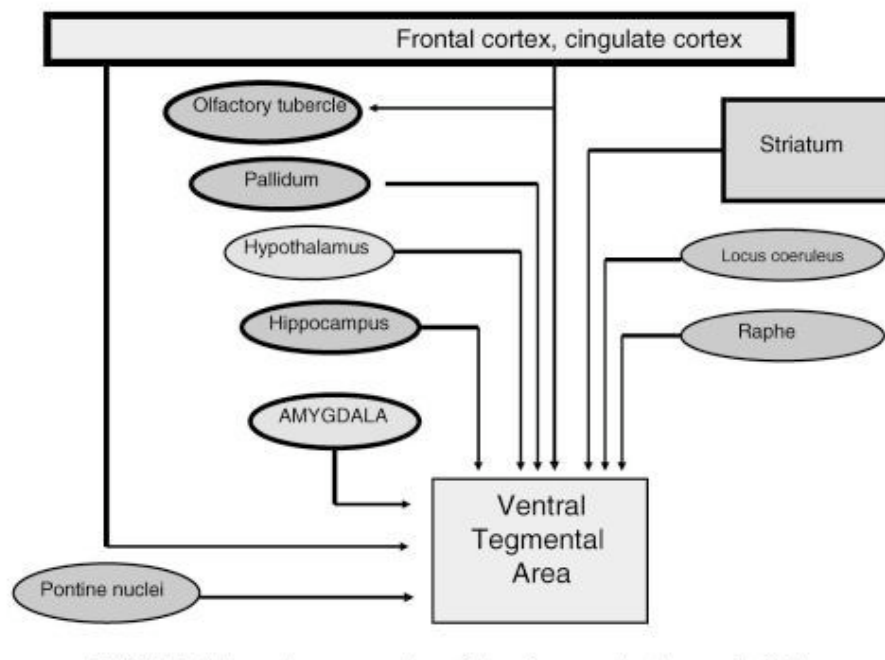of different neurotransmitter receptors on different VTA neurons determines their overall output activity.

Figure 15. VTA different afferent projections - picture taken from: Miller Cummings "The Human Frontal lobes"

These afferent imputs are possibile targets for addictive drugs, as well as for current pharmacological treatments in psychiatric disorders, including schizophrenia and depression.

The PFC plays an important role in the neurobiology of "attention", in the choice of appropriate responses, spatial learning, temporal planning and sequencing of motor actions, planning of future actions based on prevoiusly acquired information, and working memory (Castner Goldman-Rakic and Williams 2004, Goldman-Rakic et al 2004). Disruption of DA receptor in the dorsolateral PFC is associated with profound working memory alterations, in agreement with a DA receptor mulfunction in schizophrenia (Castner et al 2004), This is particularly relevant since a working memory dysfunction is the most prominent deficit in patients with schizophrenia (Park et al 1999). Notably it has been observed that DA receptor antagonists alleviate the symptoms of schizophrenia (Seeman 2002) whereas elevated DA levels in PFC correlate with a more severe impairment (Angrist et al 1985). In particular D1 type of DA receptors are the most promising target for

treatment of PFC cognitive dysfunctions (Castner et al 2004).

Increased dopaminergic activity in the mesolimbic pathway is consistently found in schizophrenic individuals. Schizophrenia is a psychosis, a disorder of thought and sense of self. Although it affects emotions, it is distinguished from mood disorders in which such disturbances are primary. Similarly, there may be mild impairment of cognitive function. However, it is distinguished from the dementias in which disturbed cognitive function is considered primary. There is no characteristic pathology, such as neurofibrillary tangles seen in Alzheimer disease. Schizophrenia is a common disorder with a lifetime prevalence of approximately 1%. It is highly heritable but the genetics are complex. PET studies suggest that the less the frontal lobes are activated during working memory task, the greater the increase in abnormal dopamine acivity in the striatum, thought to be related to the neurocognitive deficits in schizophrenia (Meyer-Lindenberg et al 2002). VTA neurons have been also strongly implicated in development of drug addiction behaviours.

The mesolimbic pathway has been implicated in tranlation of motivation to motor behaviour- and reward-related learning. The most consistent part of VTA involved in these tasks establish its connections with the nucleus accumbens in ventral striatum.

This system is commonly implicated in the seeking out and consumption of rewarding stimuli or events, such as sweet-tasting foods or sexual interaction. However, its importance to addiction research goes beyond its role in "natural" motivation: while the specific site or mechanism of action may differ, all known drugs of abuse have the common effect since elevate the level of dopamine in the nucleus accumbens. This may happen directly, such as through blockade of the dopamine re-uptake mechanism (cocaine), or indirectly, such as through stimulation of the dopamine-containing neurons of the VTA that form synapses onto neurons in the accumbens (opiates). The euphoric effects of drugs of abuse are thought to be a direct result of the acute increase in accumbal dopamine (Wise 1996). Chronic elevation of dopamine will result in a decrease in the number of DA receptors available in a process known as "downregulation".

The decreased number of receptors changes the permeability of the cell membrane located post-

synaptically, such that the post-synaptic neuron is less excitable. It is hypothesized that this dulling of the responsiveness of the brain's reward pathways contributes to the inability to feel pleasure, known as anhedonia, often observed in addicts. The increased requirement for dopamine to maintain the same electrical activity is the basis of both physiological tolerance and withdrawal associated with addiction. At the molecular level the drug addiction has been related to a change in the synaptic plasticity which normally takes place in learning processes.

Parkinson`s disease

Parkinson`s disease: retrospective, clinical features and epidemiology

Parkinson`s disease was described for the first time by an english physician James Parkinson, in his essay "An Essay on the shacking Palsy" (1817). It is a clinical syndrome, with variable combinations of akinesia, rigidity, resting tremor and postural instability and a characteristic pattern of neurodegeneration, which predominantly affects dopaminergic neurons of the substantia nigra, leading to a depletion of dopamine in their striatal projection areas (Bernheimer et al 1973, Selby 1984, Trétiakoff 1919). The prevalence of Parkinson's disease is approximately 1-2% worldwide with a typical age-of-onset of about 60 years (Shastry 2001). The characteristic symptoms appear first when about 50% of the dopamine neurons in the SN are lost, and the levels of dopamine in the striatum are reduced by 80% (Agid 1991, Bernheimer et al 1973, Kish et al 1988). Patients suffering from Parkinson's disease often display reduced spontaneity, lack of motivation as well as dementia (Braak and Braak 2000, Jellinger 1991, Zweig et al 1993).

The annual incidence of PD ranges between 16 and 19 individuals per 100,000 (Twelves et al 2003). PD occurs throughout the world in all ethnic groups and affect both sexes roughly equally or with slight predominance among males (Zhang and Roman 1993). In 5-10% of the patients initial symptoms arise between age 10 and 40 years, hence is defined as young onset PD (Golbe 1991). However, the first symptoms of juvenile onset PD occur before the age of 20 (Muthane et al 1994).

Young onset and juvenile onset PD are familial PD since family history is found in these patients. Idiopathic PD or typical sporadic PD is the term probably reserved for PD cases without or known etiology or familial connection (Samii et al 2004).

Another hallmark of Parkinson's disease is the presence of intraneuronal cytoplasmatic inclusion bodies, Lewy bodies. These eosinophilic protein aggregates composed of specific cytoplasmatic proteins, such as α-synuclein, parkin, ubiquitin, synphilin and oxidized neurofilaments, were first described in 1912 (Lewy 1912, Forno et al 1996).

Lewy bodies are primarily, but not exclusively, found in the SN. However, the presence of Lewy bodies is not exclusive for Parkinson's disease. A group of disorders which are called Parkinson-plus diseases, display the classical features of Parkinson`s disease with additional symptoms that distinguish them from the simple idiophatic PD. These include: multiple system atrophy (MSA), progressive supranuclear palsy (PSP) corticobasal degeneration (CBD), and dementia with Lewy bodies (DLB).

All of these disorders have in common, the presence of Lewy body-like inclusions and for this reason they have been also known as alpha-synucleinopathies due to the fact that alpha-synuclein is a major component of the inclusions (Spillantini 1999, Iwatsubo 2007).

In DLB cases, the distribution of Lewy bodies is denser and more widespread than in idiophatic PD. Lewy bodies are also present in Alzheimer's disease, as well as in healthy people during normal aging (Forno et al 1996, Gibb and Lees 1988, Spillantini et al 1998). However, the exact role for Lewy bodies in Parkinson's disease is not established (reviewed by Harrower et al 2005, Halliday and McCann 2008). Neuromelanin, a polymer pigment, is present in the SN (black nuclei) and the LC (blue spot) and increases with age (Marsden 1983). The exact function is not known, but it has been suggested to be a waste product in the metabolism of catecholamines. Extracellular neuromelanin in the SN, however, is found to correlate to areas within the SN high in immunoreactive MHC class II microglia (Beach et al 2007). Additionally, neuromelamin can stimulate the release of tumor necrosis factor alpha (TNFα), interleukin- 6, and nitric oxide from

cultivated microglia (Wilms et al 2003). The neuromelanin-pigmented neurons are the neurons undergoing degeneration in Parkinson's disease. The levels of neuromelanin are about half in brains from Parkinson's disease patients compared to aged matched controls (Zecca et al 2002). This loss of neuromelanin in Parkinson's disease can be monitored using magnetic resonance imaging (MRI) (Sasaki et al 2006, reviewed by Zecca et al 2006).

In addition to the loss of VM dopamine neurons in Parkinson's disease, the noradrenaline neurons in the LC (A7) are also degenerating (Ehringer and Hornykiewicz 1960). The loss of neurons in the LC is actually more extensive than that found in the VM (Ehringer and Hornykiewicz 1960, German et al 1992, Mann 1983, Zarow et al 2003). It has been demonstrated that noradrenaline reduces oxidative stress in VM cultures (Troadec et al 2001). Furthermore, the loss of noradrenergic neurons results in increased sensitivity in dopamine neurons to various insults, e.g. 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) (Heneka et al 2003, Mavridis et al 1991, Srinivasan and Schmidt 2004).

Parkinson`s disease: Genetics

Genetic discoveries in combination with epidemiologic studies have implied multiple putative causes for Parkinson's disease, thus the term "Parkinson's diseases" might be more appropriate (Galpern and Lang 2006, Lang 2007).

Some of the features implicated to participate in the destruction of the dopaminergic neurons are age, genetic factors, environmental factors, neuroinflammation, and oxidative stress. Among these, age is the strongest risk factor for developing the disease (Semchuk et al 1993).

Results of twin study, a frequently used tool to study the role of genetics, showed little difference in concordance in twins when PD develops after age of 50 years, whereas complete concordance was shown in monozygotic twins for PD with onset before 50 years of age. This finding suggests that genetic factors play a more significant role in the young-onset than late-onset or typical sporadic PD

(Tanner et al 1999). In the population of PD patients there were less than 5% patients with monogenetic inheritance of the disease, 10-15% patient with familial history and vast majority of patients, about 80%, were essentially sporadic or idiopathic with complex etiology (Lang and Lozano 1998, Olanow and Tatton 1999). Current information suggests that PD is likely to result from a combination of both genetic and environmental factors (Olanow and Tattonc 1999, Hardy et al 2003). Several genes have been suggested to participate in either dominant or recessive forms of PD with different age at onset and different associated features.

These genes can be grouped in several functional categories: proteins involved in organelle trafficking and vescicular fusion (α-synuclein, tau), genes implicated in degradation pathways like ubiquitin-proteasomal mediated degradation (parkin, UchL1, DJ-1) and lysosomal function (β-glucocerebrosidase), genes affecting mitochondria (PINK1, DJ-1, LRRK2, HtrA2, PolG), proteins that modify oxidative stress response or antioxidant functions (sepiapterin, DJ-1, Fgf20) (Funayama et al 2002, Maroteaux et al 1988, Nagakubo et al 1997, Shimura et al 2000, Valente et al 2004, Zimprich et al 2004, reviewed in Gasser 2007, Inamdar et al 2007) (Table 3).

PARK designation / OMIM accession	Gene	Mode of inheritance	Initial locus identification	Currently known prevalence	Suggested normal function
Park1 - 168601	alpha-synuclein	Autosomal dominant	Polymeropoulos, M.H. et al. (1997)	13 families from Italy, Germany, Greece	Inhibits synaptic vescicle priming Larsen, K.E. et al. (2006)
Park2 - 600116	Parkin	Autosomal recessive in early onset, susceptility in adult	Kitada, T. et al. (1998) Zhang, Y. et al. (2000)	Commonly associated to juvenile PD cases	Ubiquitin E3 ligase may be involved in Lewy body formation since absent in juvenile PD
Park3 - 602404	2p13 strongly suspected to be sepiapterin	Autosomal dominant	Sepiapterin mutation causes a DOPA-responsive dystonia	Six families from Denmark and Germany	Sepiapterin synthesizes tetrahydrobiopterin factor
Park4 - 163890	alpha-synuclein duplications	Autosomal dominant	Singleton, A.B. et al. (2003)	Seven families in Europe, USA and Japan	Inhibits synaptic vescicle priming Larsen, K.E. et al. (2006)
Park5 - 191342	UchL1	Possibly autosomal dominant	Leroy, E. et al. (1998)	Single sibling pair in Germany	Hydrolizes polyubiquitin
Park6 - 605909	PINK1	Autosomal recessive in early onset	Valente, E.M. et al. (2004)	Three related families in Sicily	Mitochondrial serine/ threonine kinase Beilina, A. et al. (2005)
Park7 - 606324	DJ-1	Autosomal recessive	Bonifati, V. et al. (2003)	Families in Holland, Italy and Uruguay	Sumoylation pathway, antioxidant, Martinat, C. et al. (2004)

Park8 - 609007- 607060	LRRK2 / dardarin	Autosomal dominant	Paisan-Ruiz, C. et al. (2004) Zimprich, A. et al. (2004)	Very common in North Africa and middeastern populations	Kinase with a GTPase activity, Smith, W.W. et al. (2005) West, A.B. et al. (2005)
Park9 - 610513- 606693	ATP13A2	Autosomal recessive	Ramirez, A. et al. (2006)	One jordanian and one chilean family	Lysosomal transporter and ATPase Schultheis, P.J. et al. (2004)
Park10 - 606852	Unknown 1p32	Autosomal susceptibility factor which influences age of onset	Not identified	Based on large population studies in Iceland - unclear	Unknown
Park11 - 607688	Unknown 2q36-p37	Autosomal dominant	Not identified	Identified in sib pair, might be common in familial PD	Unknown
Park12 - 300557	Unknown Xq21-q25	X-linked	Not identified	Unclear	Unknown
Park13 - 610297	Omi/HtrA2 serine protease-25	Autosomal dominant	Strauss, K.M. et al. (2005)	In four patients in Germany	Mitochondrial serine protease
230800	Beta-glucocerebrosidase	Autosomal susceptibility, recessive for Gaucher`s	Associated with Gaucher disease type 1, the most common lysosomal storage disorder	many families particularly Askenazi families	Hydrolizes glucoceramide within lysosomal degradation

Table 3. Genes determined-suggested to be cause or suceptibility factors for primary parkinsonism.

Parkinson`s disease: alpha-synuclein

The A30P and A53T α-synuclein (α-syn) mutations were the first genes found to be linked to familial PD (Polymeropoulos et al 1997) and even before α-syn was detected as major component of Lewy bodies in PD, a role of α-syn in AD was discovered, when a peptide derived from the central hydrophobic region of this protein (aa 61-95), known as non-amyloid beta component of AD amyloid plaques (NAC), was found to represent the second major instrinsic element of the AD senile plaques (Ueda et al 1993, Han et al 1995). The Snca gene encodes for a small acidic protein first described in *Torpedo californica* and very abundant in brain (Clayton and George 1998, Tofaris and Spillantini 2005), which is highly expressed in neurons and glial cells of the peripheral and central nervous system, where it is primarily concentrated in presynaptic axon terminals. It belongs to the synuclein family which comprises β-synuclein and γ–synuclein sharing a common signature sequence in their N-terminal region containing a different number of repat regions while they differ in their carboxy-terminal part (Tofaris and Spillantini 2005).

α-synucleins from different organisms possess a high degree of sequence conservation (Clayton and George 1998). At least three α-synuclein isoforms are produced in humans by alternative splicing (Beyer 2006). The best-known isoform is α-syn-140 (NACP140), which is the whole and major

transcript of the protein. Two other isoforms, α-syn-126 and α-syn-112 (NACP112), are produced by alternative splicing, resulting in an inframe skipping of exons 3 and 5, respectively. Exon 3 encodes for residues towards the N-terminus, residues 41–54, and exon 5 encodes for residues towards the C-terminus, residues 103–130. The whole transcript, α-syn-140, can be divided into three regions:

- The N-terminal region, residues 1–60, includes the sites of three familial PD mutations and contains four 11-amino acid, imperfect repeats with a highly conserved hexameric motif (KTKEGV). The N-terminal region is positively charged and predicted to form amphipathic α-helices, typical of the lipid-binding domain of apolipoproteins, with which α-syn shares homology (George et al 1995, Clayton and George 1998).

- The central region, residues 61–95, comprises the highly aggregation-prone NAC sequence (Ueda et al 1993, Han et al 1995).

- The C-terminal region, residues 96–149, is highly enriched in acidic residues and prolines. Three highly conserved tyrosine residues, which are considered a signature of the α- and β-synuclein family, are located in this region.
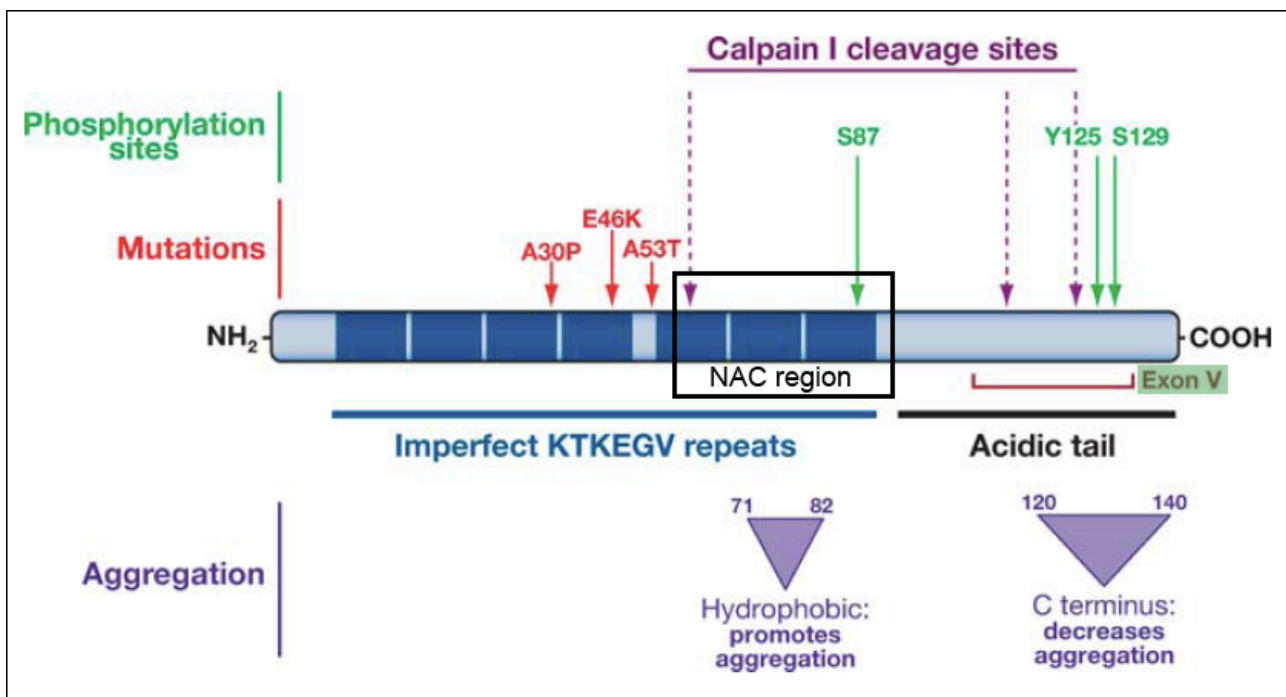

Figure 16. Motifs in the α-synuclein protein. The natively unfolded α-synuclein protein is shown in a linear form. PINK1 shaded areas represent the imperfect KTKEGV repeats. Human mutations are shown in red and map to the repeat region. At the C-terminal end of the protein is an acidic tail, containing several sites of phosphorylation (*green*). The C-terminus also contains the alternatively spliced exon V and a calpain I cleavage site. The acidic tail tends to decrease protein aggregation, whereas a hydrophobic region near the imperfect repeats promotes aggregation. Modified from (Cookson 2005).

The α-syn belongs to a family of natively unfolded proteins that lacks a typical secondary structure. Under pathological conditions, these disordered proteins can also undergo conformational changes leading to protein aggregation and deposition. In the case of α-syn, its pathological fibrillization results in the formation of intracellular inclusions and neurodegeneration. Different functions have been associated to the three main domains in the protein (Figure 16). The seven imperfect 11-residue repeat sequences are predicted to form five amphipathic helices on the amino-terminal half (Bisaglia et al 2006, Davidson et al 1998, Jao et al 2004). On the other hand, the acidic, glutamate-rich carboxy-terminal region remains unstructured even in the presence of membranes (Jao et al 2004). α-syn helices 1–4 are predicted to associate with lipid vesicles (Biere et al 2000, Perrin et al 2000), while helix 5 would be responsible for protein–protein interactions (Davidson et al 1998). Nevertheless, α-syn is a very dynamic molecule whose secondary structure depends on its environment (Davidson et al 1998, Perrin et al 2000). Thus, it adopts an unfolded random coil structure in aqueous solution (Weinreb et al 1996) and shows an α-helical structure upon binding to acidic phospholipid vesicles (Bussell and Eliezer 2004, Chandra et al 2003, Trojanowski and Lee 2003). Finally, as with other amyloidogenic proteins, α-syn undergoes a pathological transition from random coil to a β-pleated sheet conformation accompanied by extensive aggregation and fibril formation resembling the filaments of Lewy bodies (Conway et al 2000, Giasson et al 1999, Serpell et al 2000).
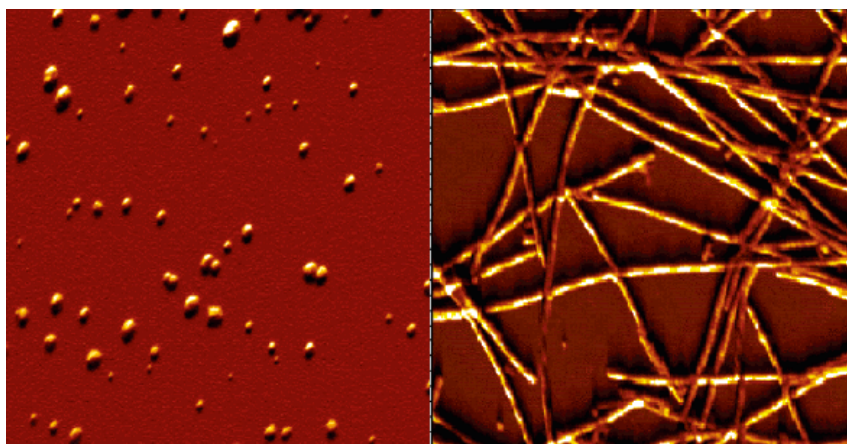

Figure 17. AFM images of R-synuclein fibrils (right) and oligomers (left). The fibrils show the characteristic twist of mature fibrils and are 10 nm high. The oligomers are from the late lag phase period and are mostly 5 or 7 nm high. The images are 3 μm square. Modified from Fink Acc. Chem. Res. **2006,** 39, 628-634

α-Syn exists as two different structural conformers *in vivo*, one of which is a membrane-bound form with a stable α-helical structure and the other an essentially disordered free cytosolic form (Bussell et al 2005, Eliezer et al 2001, Jo et al 2000). It has been shown that these membrane-bound α-syn forms initiate aggregation and provide the seeds responsible for accelerated deposition of the less aggregation prone but more abundant cytosolic form (figure 17).

Some hints about the physiological function of α-syn comes both from the analysis of the structure of the protein, its cellular localization and the phenotype associated to its mutations.

The biochemical structure of the protein predicts that α-syn may function as a molecular chaperone capable of binding to other intracellular proteins (Ostrerova et al 1999). This hypothesis was strengthened by two observations. First, α-syn shares structural homology with the 14-3-3 family of molecular chaperone proteins (Ostrerova et al 1999, Tzivion et al 1998).

Second, α-syn expression is regulated throughout development, with up-regulation during periods of accelerated neuronal activity and plasticity (Petersen et al 1999, George te al 1995).

Consistent with the chaperone hypothesis, α-syn interacts at axon terminals with numerous proteins, more than 250, some of which regulate dopamine homeostasis in nigrostratial dopamine neurons (Zhou et al 2004). One of the first identified interactors is phospholipase D2, which may regulate monoaminergic vesicle content and nerve terminal dopamine storage (Jenco et al 1998, Lotharius and Brundin 2002). Additionally, phospholipase D2 may be involved in cellular adhesion (Powner et al 2005) and the recycling pathway between endosomes and the plasma membrane (Padron et al 2006).

α-syn expression also alters dopamine transporter- mediated uptake of synaptic dopamine, which suggests a role in terminating dopamine neurotransmission (Sidhu et al 2004). The catalytic activity of tyrosine hydroxylase, the rate-limiting enzyme in dopamine synthesis, regulates α-syn through phosphorylation (Perez et al 2002, Drolet et al 2006). Furthermore, α-syn plays a key role in mediating substrate-dependent enhancement of phosphorylated tyrosine hydroxylase and dopamine

synthesis in nigrostratial dopamine neurons (Perez et al 2002, Drolet et al 2006) and is directly involved in dopamine recruitment and presynaptic dopamine compartimentalization (Yavich et al 2004). Recently, it has been shown that aggregated α-syn stimulates the phosphorylation of TH (Alerte et al 2008).

Finally, an antiapoptotic function of α-syn by down-regulation of the p53 pathway has also been observed (Alves Da Costa et al 2002).

Among the pathogenic genes implicated in PD, α-syn is probably the most studied since:

- _ It is a major component of LBs and Lewy neurites (Spillantini et al 1997).

- _ Overexpression of wild-type α-syn through gene multiplication (also known as the PARK4 gene for familial PD) triggers PD (Singleton et al 2003).

- _ Virus-mediated α-syn overexpression can kill SN DA neurons (Kirik et al 2002).

- _ α-syn knockout mice are resistant to DA neuronal death by MPTP toxicity (Dauer and Przeborski 2003).

- _ MPTP and 6-hydroxy-DA increase α-syn expression in rodent SN DA neurons (Vila et al 2000, Kholodilov et al 1999).

PD is now widely considered a 'synucleinopathy' as are 'parkinson-plus' disorders such as multiple system atrophy (MSA) and diffuse Lewy body disease. These diseases also show accumulations of aggregated α-syn. The physiological role of α-syn might involve the control of vesicle exocytosis at a late stage before fusion (Larsen et al 2006). This might involve alternate binding and dissociation from membrane, shifting between an α–helical conformation in 'lipid rafts' to an unfolded conformation in the cytosol. Provocatively, the A30P PD mutation binds the membrane poorly and might be more subject to aggregation. The dissociation from the synaptic vesicle membrane seems to result from and regulate synaptic vesicle exocytosis (Fortin et al 2005). It is by no means clear whether the normal physiological function of α-syn is involved in PD, and it is widely suspected that problems in degradation might initiate a 'toxic gain of function', perhaps through aggregation.

Braak et al. (2006) examined immunolabeled α-syn to study aggregation in autopsy of PD patients and also in asymptomatic individuals who might have been en route to developing PD. They suggest a progression of PD through ascending brain regions, beginning with medulla oblongata/pontine tegmentum and olfactory bulb, to locus coeruleus (LC), raphe, then SN, and at late stages to the neocortex. There is no clear relationship between α-syn expression and neuronal cell death. Expression is much higher during early human development than at maturity (Raghavan et al 2004), when pathogenesis takes place. However α-syn expression has been shown to increase progressively during normal ageing in monkeys and humans (Chu and Kordower 2007).

In normal adult humans, the α-syn mRNA label is strong in neuromelanin (NM) positive SN neurons but is also high in unaffected regions (Kingsbury et al 2004). Moreover, there are no data indicating extensive neuronal loss of cortical neurons in PD despite the presence of LBs. In other words, the presence of α-syn pathology in neurons does not indicate that the neuron will necessarily die.

Not surprisingly, products resulting from the interaction between DA and α-syn have been extensively explored for toxicity. Originally presumed to be a covalent bond, recent results suggest that there is an ionic interaction between DA–quinone and residues 125–129 in the protein (Norris et al 2005). DA modification of α-syn seems to maintain small α-syn oligomers in a reactive protofibril conformation by inhibiting progression to a less reactive aggregate. Protofibrils damage the membrane and do so more effectively with the pathogenic α-syn mutations (Volles et al 2001).

Micromolar concentrations of α-syn cause proton leakage from isolated chromaffin secretory vesicles (Mosharov et al 2006), and α-syn molecules with pathogenic mutations are particularly effective at causing leakage. It is not known whether the leak is caused by protofibril or native forms, as α-syn is amphiphilic and might have detergent-like properties.

If DA synaptic vesicles are damaged by α-syn, a vicious cycle of dysregulated cytosolic DA and further α-syn damage may occur. DA-modified α-syn has been shown to block CMA in a manner

similar to that of the pathogenic mutants (Martinez-Vincente et al 2008). As a consequence, when neurons are stressed, CMA could be induced as a protective pathway, but for SN neurons with reactive levels of DA, the resulting α-syn modification could block CMA and lead to specific DA neuronal death.

Parkinson`s disease: multiple transcripts of Snca

At least four different alternative splicing variants of the α-syn transcript have been reported to exist in human (Beyer 2006). α-syn 140 is encoded by the larger transcript, whereas alternative splicing of exons 3 and 5 gives rise to α-syn 126 and α-syn 112, respectively. As shown in figure 18, both α-syn 126 and α-syn 112 are characterized by the in-frame deletion of the correspondent exon sequence with no premature stop codons (Beyer et al 2006). More recently, another splicing isoform resulting from the in-fame deletion of both exon 3 and 5, called a-syn-98, has been described (Beyer et al 2008). Since alternative splicing generates potentially functional proteins with a different combination of domains, this has been believed it could be of particolar relevance for studying the functional roles associated to each specific isoform.

α-syn 140 is a small protein with a molecular weight of 14.5 kDa. The isoforms generated by alternative splicing show small molecular weight differences (13.1 kDa for α-syn 126 and 11.4 kDa for α-syn 112) and, therefore, their SDS-page separation and Western blot detection are difficult to carry out, particularly when trying to distinguish between α-syn 140 and α-syn 126. Although Li et al analyzed C-terminal truncated α-syn forms, in their study the possibile existence of α-syn isoforms was not taken into account (Li et al 2004). The present lack of isoform-specific antibodies hampers the detection of α-syn isoforms at the protein level. Nevertheless, there exists some in direct proof that at least α-syn 112 is expressed as a protein. Studies on early α-syn accumulation preceding Lewy body formation (Kuusisto et al 2003) and on membrane bound α-syn (Lee et al

2002) failed to detect α-syn with antibody LB509. This antibody is commonly used in studies of α-synucleinopathies and recognizes a minimum epitope at residues 115–122 of α-syn (Jakes et al 1999), which is localized within the amino acid sequence corresponding to Snca gene exon 5 (this exon is lacking in α-syn 112). In contrast, in both studies α-syn detection was carried out with the aid of antibodies recognizing the C-terminal part (Jakes et al 1999, Lee et al 2002).
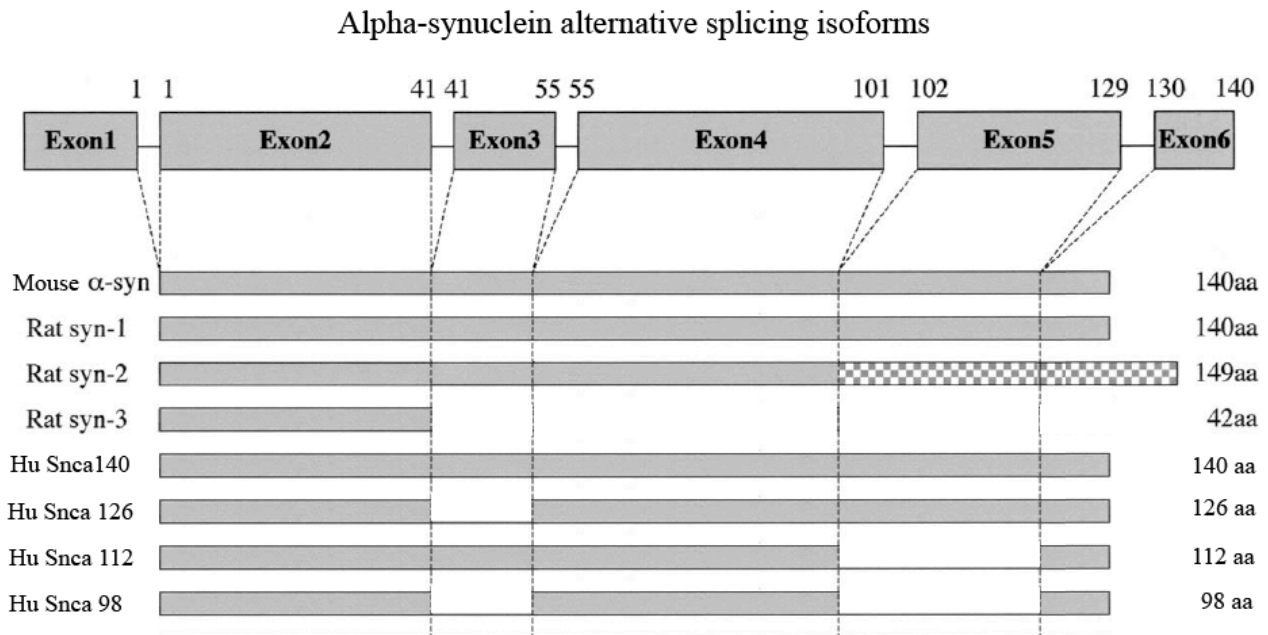

Figure 18. Comparison of the alternative splicing isoforms of the α-synuclein gene that were experimentally determined; the four different human splicing isoforms (Snca 140, Snca 126, Snca 112, Snca 98) were confirmed to be expressed in human brain (Beyer et al 2008, Beyer 2006); whereas three splicing variants are described in rat (MacLean and Hyman 2002).

α-Syn 140 is preponderantly transcribed in higher levels than α-syn 126 and α-syn 112 (Campion et al 1995). As α-syn 140 is the whole transcript of the Snca gene, it conserves its whole structure as well as all posttranslational modification sites. It is worth mentioning that all studies of α-syn structure and function have been carried out considering the whole protein only.

Thus, although several α-syn functions have been identified in the last few years, nothing is known about the specific roles of each of the three α-syn isoforms (Beyer et al 2004). α-syn 126 is characterized by the in-frame deletion of the sequence corresponding to exon 3 (Figure 18). Exon 3 localizes at the N-terminal of the protein and encompasses amino acid residues 41–54. The splice-

out of exon 3 results in the interruption of the four helical protein–membrane interacting domain with the deletion of almost the entirety of helix 3 and part of helix 4 (Figure 18). Interestingly, two of the three α-syn mutations described up to now are located within the sequence corresponding to exon 3. Specifically, the E46K mutation is in the midportion and the A53T mutation at the end of exon 3 (Polymeropoulos et al 1997, Zarranz et al 2004). The importance of α-syn N-terminal half during aggregation was suggested by α-syn intrinsic structure (Bertoncini et al 2005, Bisaglia et al 2006b). Fibrillation studies of A53T and E46K mutated α-syn further underscored the importance of α-syn N-terminus in modulating filament formation (Greenbaum et al 2005).

Additionally, mutant E46K α-syn showed increased formation of α-syn aggregates when compared to cells expressing wild-type or mutant A53T or A30P α-syn (Narhi et al 1999, Pandey et al 2006). Since structurally the E46K mutation affects the fourth KTKEGV repeat, this sequence seems to be of special importance in α syn aggregation. Interestingly, important differences in α-syn 126 brain expression levels between the three neurodegenerative disorders and controls could be observed (Beyer et al 2006). The expression of α-syn 126 was markedly decreased in the cortex in three neurodegenerative disorders DLB, LBVAD, and AD, with an almost fivefold decrease in DLB when compared to controls (Beyer et al 2006).

α-syn 112 lacks the sequence corresponding to exon 5, located on the C-terminal half of the protein (Figure 18). The result of this in-frame deletion is a 112 amino acid protein lacking amino acids 103–130. It has been shown that α-syn is phosphorylated at serine 129. Although α-syn 112 lacks this site, a second α-syn phosphorylation site at serine 87 would remain unaltered in this isoform (Okochi et al 2000). So far no studies have been carried out to determine the consequences of α-syn 112 phosphorylation at serine 87. However, it has to be noted that serine 87 is not conserved across species, and is absent in mouse and rat. Finally α-syn 98 lack both exon 3 and exon 5 and has been shown to be expressed at variable levels in various areas of the human brain (Beyer et al 2008). Overexpression of α-syn 98 in LBD and AD patients compared to control subjects has been

reported for this short isoform, suggesting that it can play important roles in the pahogenesis of these disorders (Beyer et al 2008).

In mouse so far there is no experimental evidence of expression of alternative splicing isoforms of α-syn, but many gene structure prediction tools report some different transcript isoforms produced from the Snca locus. Here you can find Aceview annotations summarizing what is known about the possible mouse transcript isoform based on the EST and cDNA data from various databases.
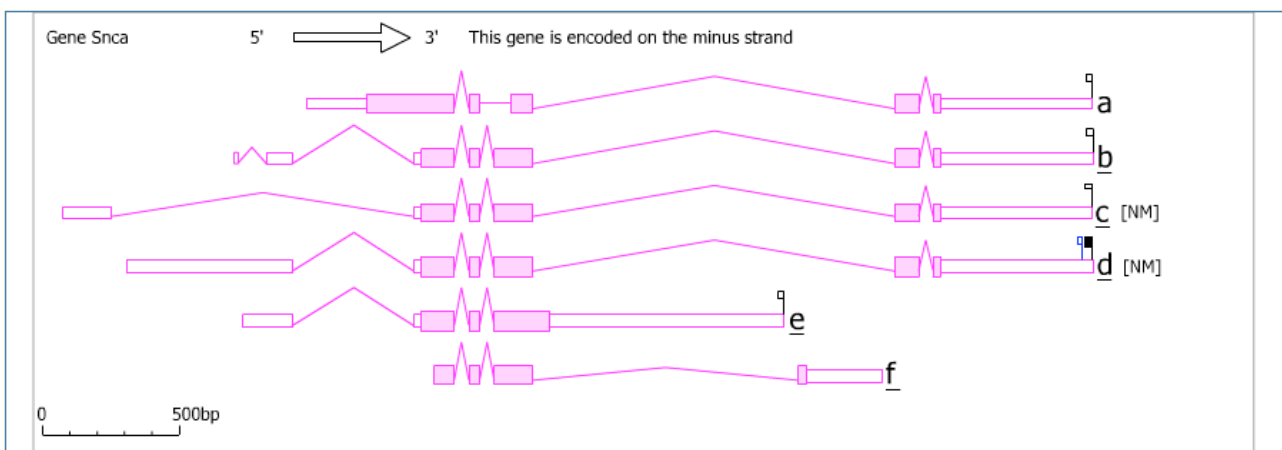
Protein	Protein quality	Exons in CDS	Domains	Predicted localization	Completeness and uniqueness	Extends from	coordinates on mRNA	minimal set of supporting clones
.a	186 aa Very good	5	Synuclein		complete	Met (a..CTG.) to Stop	223 to 783	AW988661 CB173579
.b	140 aa Very good	5	Synuclein	cytoplasmic	complete =cSep07	Met (ATG) to Stop	135 to 557	AF033261
.c	140 aa Very good	5	Synuclein	cytoplasmic	complete	Met (ATG) to Stop	212 to 634	AF179273
.d	140 aa Very good	5	Synuclein	cytoplasmic	complete =cSep07	Met (ATG) to Stop	637 to 1059	AK156316
.e	122 aa Very good	3	Synuclein	cytoplasmic	complete	Met (ATG) to Stop	208 to 576	AK014472
.f	96 aa Very good	4	Synuclein		COOH complete	1st codon to Stop	1 to 291	AA242305

Figure 19. A scheme of the possibile alternative transcripts and their corresponding proteins associated to the mouse Snca locus as emerges from cDNA and EST data collected from many tissues and summarized by the AceView gene prediction database. At least 6 different isoforms are reported for α-syn; the d isoform being the most represented one with 180 accessions and being expressed in brain, visual cortex, ofactory brain, embryo. The other isoforms are less abundant: transcript e (80 accessions from visual cortex, whole brain, lateral ventricle wall, liver, activated spleen); transcript c (17 accessions from brain, embryo, prostate and lung); transcript b (3 accessions from retina); transcript a (2 accessions from mammari gland, offactory and respiratory epithelium); and transcript f (1 accession from fethus).

On the other hand in rat, three alternative splicing isoforms for α-syn have been described (reported in figure 18). Interestingly, one of these isoforms, rat α-syn 2 is expressed *in vivo* in adult brain and it has been observed to accunulate in cytoplasmic spots when overexpressed in cultured transfected

H4 cells. Morover, the treatment with the proteasome inhibitor *n*-acetyl-leu-leu-norleucinal (ALLN), induces the α-syn 2 isoform to specifically forms aggregates that were not seen for rat α-syn 1. The two isoforms differ in their C-terminal region where a different sequence of 48 aa, specifically found in α-syn 2, is absent in α-syn 1 (McLean and Hyman 2002).

Parkinson`s disease: molecular pathways

Although various groups of neurons exhibit pathology in PD, the parkinsonism is clearly caused by loss of DA following death of substantia nigra (SN) neurons, as shown by the extensive (about 80%) loss of neuromelanin-pigmented (NM+) DA neurons in the SN with a comparative sparing of the neighboring NM- ventral tegmental area (VTA) DA neurons (German et al 1989, Hirsh et al 1988). PD pathogenesis in nearly all cases converges upstream of LB formation and death of DA SN neurons. Some hints into the mechanisms for specific vulnerability in PD would come from a deep knowledge and understaning of the biology of the SN and VTA neurons, and what is unusual about the A9 cells in particular. It has been suggested that those neurons might have features that predispose them to be extremely susceptible to stress.

Among the multiple events that can concurrently lead to Parkinson`s disease there are factors involved in many different pathways and functions. The link between Parkinson's disease and mitochondria was first established with the identification of a deficiency in the activity of complex I (NADH ubiquinone reductase) in PD substantia nigra and subsequently in the peripheral tissues of patients. Importantly, inhibition of complex I resulted in increased free radical generation and could contribute to the oxidative damage seen in SN. Inhibition of mitochondrial respiration reduced the levels of ATP required for normal cellular function.

These changes in turn may induce protein damage and misfolding (Cassarino and Benneltt 1999). Neurotoxic compounds as MPTP and rotenone, which selectively damage dopaminergic neurons and cause PD-like clinical features, are selective inhibitors of complex I of mitochondrial

respiratory chain (Greenamyre et al 2003). There is 30-40% decrease in complex I dysfunction in PD which could thus play a role in oxidative stress, free radical-mediated protein damage, impaired proteasomal function and protein accumulation (Shamoto-Nagai et al 2003). Mitochondria are also intimately involved with both necrotic and apoptotic cell death. Recently, gene mutations of mitochondrial proteins PINK1 and DJ1 caused Parkinson's disease, and mutations of MtDNA polymerase (PolG) caused parkinsonism. The relation between mitochondrial dysfunction and PD has been highlighted further by the recent description of mtDNA abnormalities in PD patients. Occasional mtDNA point mutations have been identified in PD but these have not been present in the general PD population.

A mutation in the mtDNA 12S RNA was identified in a patient with maternally inherited early onset PD, deafness, and neuropathy. Several studies that have sequenced mtDNA in PD patients have not identified any consistent mutations. Two studies have shown a link between mtDNA haplotypes and the risk for developing PD. The first showed a reduced risk for PD in individuals with haplotypes J and K and the second a 22% decrease in PD risk in those with the UKJT haplotype cluster. In contrast, a smaller study reported an increased risk for PD with haplotypes J and T.

Futhermore, studies in animal model systems of PD gained more insights into the mitochondrial contribution to PD: PINK1 and Parkin have been implicated to act in common in a pathway to control mitochondrial morphology (Poole et al 2008). Pink1 in Drosophila is required for mitochondrial function and interacts genetically with Parkin (Clark et al 2006) whereas loss of function mutations of human PINK1 lead to mitochondrial pathology which can be rescued by Parkin (Exner et al 2007). Recently, much attention has been focused on the mitochondrial fission-fusion process which contributes to dynamically control the mitochondrial morphology in the cells, and seems to be compromised in many neurodegenerative conditions (Frank 2006) (Figure 20). Surprisingly a recent paper has shown evidenses supporting a role for Pink1 as a central regulator of the mitochondrial fission-fusion machinery (Yang et al 2008).
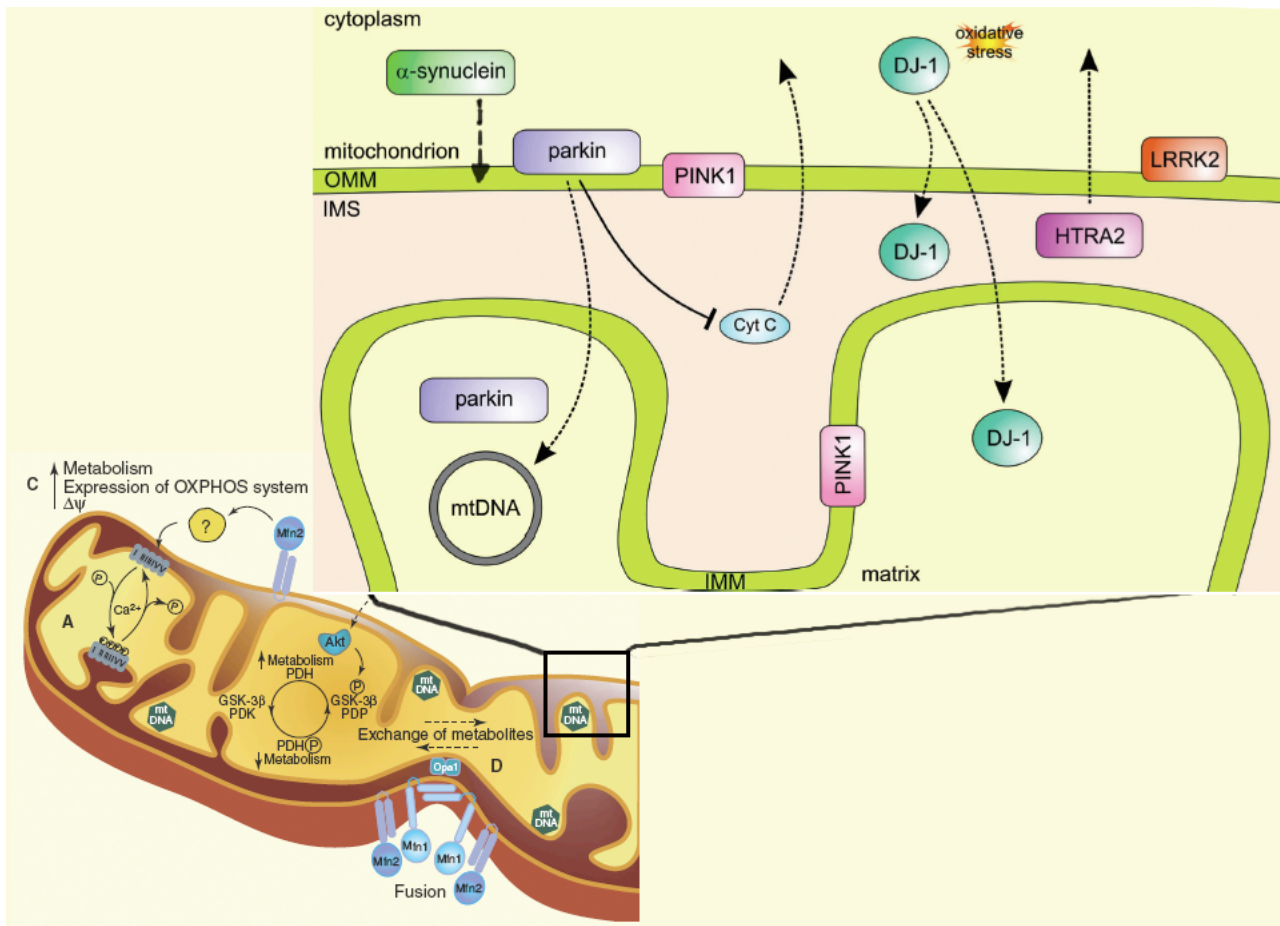
Figure 20. PD genes and their relation to mitochondria. Aggregation of mutant or overexpressed α-syn might be an upstream actor of mitochondrial alterations or with ts binding to OMM in a pH dependent manner. Parkin associates with the OMM and was shown to play a role in mitochondrial biogenesis by regulating both transcription and replication of mtDNA. PINK1 has an N-terminal mitochondrial targeting motif and is localized to mitochondrial membranes, whereas oxidation of a key Cys-residue in DJ-1 leads to its relocalization to mitochondria. LRRK2 resides diffusely throughout the cytosol, but is partly associated with the OMM. HTRA2 resides in the IMS, where from it is released upon apoptotic stimuli. Picture modified from Bogaerts et al 2008.

Oxidative stress develops from increased production of highly reactive free radicals, decreased scavenging of these reactive species, increased generation of oxidatively damaged proteins, lipids, DNA and decreased clearance of oxidized products that could be toxic for cells (Shang et al 2003). Signs of oxidative stress are abundant in the SN of patients of PD (Jenner P 2003). These include reactive oxygen species (ROS) such as hydrogen peroxide, hydroxyl, superoxide and peroxyl radicals. Other oxidants are reactive nitrogen species (RNS) as nitric oxide (NO). These strong oxidants react with DNA, protein and lipid altering their structure and causing cellular damage (Jenner 1998). The oxidants as ROS are produced by dopamine metabolism, the electron transport chain of mitochondria, iron metabolism and inflammation.

The enzymatic metabolism or auto-oxidation of dopamine generates free radicals (Hald and

Lotharius 2005). Dopamine metabolism by MAO-A leads to production of dihydroxy-phenylacetic acid (DOPAC) and H_2O_2 with the consumption of O_2 and H_2O. Intracellular auto-oxidation of dopamine generates dopamine-quinone and H_2O_2. Dopaminequinone participates in nucleophilic reaction with sulfhydryl groups of proteins leading to structural modification and reduced glutathione levels (Stokes et al 1999), and it is shown to inhibit glutamate and dopamine transporter function in synaptosome (Berman and Hastings 1997), inhibit tyrosine hydroxylase (Kuhn et al 1999) in cell free systems and promote H+ leakage from mitochondria resulting in uncoupling of respiration to ATP synthesis (Khan et al 2001). H_2O_2 produced can be converted to hydroxyl radical by Fenton reaction in presence of ferrous ion. Hydroxyl radical can react with virtually every cellular macromolecule. In the presence of ferrous ion, the superoxide anion and H_2O_2 two weakly reactive free radical species can combine in the Haben-Weiss reaction to produce the more reactive OH radical (Berg et al 2001, Hirsh and Faucheux 1998). ROS inhibit complex I and provoke mitochondrial dysfunction.

The iron-mediated catalysis of hydroxyl radical generation could be a key mechanism in pathogenesis mediated by oxidative stress. Elevation of iron levels are detected in the SNc in patients of PD (Berg et al 2001, Hirsh and Faucheux 1998). The iron levels are increased by 35% in PD patients compared to age matched control (Sofic et al 1988).

The concept of excitotoxicity has been applied to a number of neurodegenerative diseases including PD (Hallett and Standaert 2004). Glutamate is the principal excitatory neurotransmitter and mediate its action by interacting with receptors such as NMDA (N-Methyl D-Aspartate), AMPA (α-amino-3-hydroxy-5-methyl 4-isoxalopropionate) and kainate receptors. Persistent activation of the glutaminergic NMDA receptor increases intracellular levels of Ca^{2+} ions potentially leading to activation of proteases, endonucleases, phospholipases and NO synthase with resulting generation of reactive NO. Increased intracellular Ca^{2+} concentration and the subsequnt events leading to cell death is supported by the observation that dopaminergic neurons expressing calcium binding protein calbindin may be selectively preserved in PD (Hirsch et al 1992).

Parkinson`s disease: Current therapies

The most commonly medical treatment used for Parkinson's disease is L-dopa therapy. L-Dopa is a precursor of dopamine that can cross the bloodbrain- barrier. The effects of L-dopa was discovered when the impaired motor activity found after reserpin treatment was counteracted by giving the animals L-dopa (Carlsson et al 1957, Carlsson et al 1958). Further studies demonstrated dramatic reversal of the symptoms, although the effect of the drug was brief (Birkmayer and Hornykiewicz, 1962). Later, trials displayed that higher doses and oral administration of L-dopa were efficient (Cotzias et al 1967). Today, administration of L-dopa is given together with decarboxylase inhibitors to decrease the peripheral metabolism. However, after some years of treatment, the effect of L-dopa declines and the patients develop dyskinesias (involuntary movements) and/or on-off symptoms due to fluctuations in the drug dosage (Granerus 1978, Lang and Lozano 1998a, Lang and Lozano 1998b). Dopamine agonists acting on the dopamine receptors are also an important therapeutic alternative. Furthermore, treatments to hamper the breakdown of dopamine by inhibiting the catechol-O-methyltransferase (COMT) and monoamine oxidase B are undertaken. Various combinations of the above-mentioned drugs are used as therapies for Parkinson's disease. However, the medications only improve motor activity deficiencies while do not cure the disease.

One way to counteract the symptoms of the disease is to use deep brain stimulation, where high frequency stimulating electrodes are implanted in the basal ganglia. Implanting the electrodes to the ventral intermediate nucleus of the thalamus can reduce tremor (Benabid et al 1991, Putzke et al 2003). Stimulation of either the subthalamic nucleus or the globus palludus interna diminishes bradykinesia, rigidity, as well as reduces L-dopa induced dyskinesia (Benabid et al 1998, Kumar et al 1998a, Kumar et al 1998b).

Another potential approach to counteract the symptoms of Parkinsons's disease may be the replacement of the lost dopaminergic neurons. Already in 1979, fetal dopaminergic transplants were

evaluated as a putative method to counteract the symptoms of Parkinson's disease in a 6-hydroxydopamine (6-OHDA) animal model (Björklund and Stenevi 1979, Perlow et al 1979). The transplants were proven to be long-term surviving (Freed et al 1980, Stromberg and Bickford 1996), and the striatum of the host was reinnervated in a target specific manner (Strömberg et al 1992, Wictorin et al 1992). Furthermore, new synapses were formed (Freund et al 1985, Mahalik et al 1985, Stromberg et al 1988), and the grafts were able to release dopamine in the reinnervated striatum (Rose et al 1985, Stromberg et al 1988, Zetterstrom et al 1986). Ten years after the first animal transplantation, human fetal VM was grafted to patients for the first time. *Post mortem* analysis displayed graft survival and reinnervation of the striatum (Kordower et al 1998). Unfortunately, the outcome of the clinical trials revealed poor cell survival and.the striatum was not fully reinnervated (Barker et al 1996, Brundin and Bjorklund 1998, Freed et al 2001, Hagell et al 2002, Olanow et al 2003). Thus, it is important to find various neuroprotective and neurotrophic factors that can enhance survival and neuritic formation. A major problem with grafting fetal tissue is the limited availability of tissue. Stem cells, (self renewal and multipotent cells), are alternative sources for neuronal transplantation. In the last two years, much effort has been done to obtain stem cells to be used in various neurodegenerative disorders. However, major concerns regarding scientific, clinical and ethical issues need to be solved before stem cells can be used in Parkinson's disease as therapy (Newman and Bakay 2008, Parish and Arenas 2007, Wang et al 2007). Hereafter are summarized all the currently used therapies.

Neurotropic factors such as glial-derived neurotropic factors (GDNF) and brain derived neurotropic factors (BDNF) have shown potent protective and regenerative effects on dopaminergic neurons. Delivering GDNF directly into the putamen of PD patients greatly improves their symptoms and significantly increases DA uptake and storage (Gill et al 2003). BDNF appears as an important determinant of dopamine cell survival and D3 receptor expression (Guilin et al 2001). The beneficial effects of BDNF and GDNF raise the possibility that the supply of these or other

neurotropic substances may be limited or that their downstream signaling pathways may be dysfunctional in PD leading to degeneration of DA cells

Agent	Pharmacological Class	Adverse Effects
Levodopa (with carbidopa and benserazide)	Dopamine precursors	Central side effects (motor fluctuations, dyskinesia, psychiatric disturbances) and peripheral side effects (nausea, vomiting, orthostatic hypotension)
Seligiline	MAO-B inhibitors	Insomnia, neuropsychiatric side effects. Dopaminergic effects of the other drugs may be accentuated.
Tolcapone, Entacapone	COMT inhibitors	Hepatotoxicity, diarrhea, potentiation of levodopa side effects
Ergot derivatives (Bromocriptin, pergolide, lisuride, carbergoline)	Dopamine agonists	Peripheral and central dopaminergic side effects, pedal edema, pleuropulmonary reaction retroperitoneal fibrosis, erythromelagia, Raynaud's phenomenon, postural hypotension
Ropinirole, Pramipexole	D2/D3/D4 agonist	Somnolence, nausea, vomiting, dyspepsia, abdominal pain, hypotension, confusion, hallucination and dyskinesia
Apomorphine	D1/D2 agonist	Nausea and injection site reactions, neuropsychiatric side effects
Benztropine, procyclidine etc	Anticholinergics	Peripheral effects (dry mouth, blurred vision, constipation, difficulties with urination) and central effects (confusion, hallucinations, memory problems)
Amantadine	NMDA receptor antagonist	Confusion, hallucination, pedulary edema, livedo reticularis

Table 4. Currently used therapies for Parkinson`s disease – taken from Inamdar et al 2007

SECTION2. Materials and Methods

Striatal Cells and RNA extraction

For the SMART7 amplification 20 μg of total RNA was extracted from wild type and mutant striatal neurons (ST*Hdh^{Q111}/Hdh^{Q111}*), from knockin mice bearing the HD mutation (Trettel F et al, 2000; kindly provided by Persichetti F).

RNA was extracted by adding 1ml of TRIzol Reagent (Invitrogen) to two 10cm^2 Petri dishes containing approximately 10^6 cultured cells each.

For each dish, the lysate was passed several times through a pipette and incubated for 5 min at R.T. for the complete dissociation of nucleoprotein complexes. After adding 0.2ml of Chloroform and vortexing to mix, the samples were incubated at R.T. for 10 min and centrifuged at 12,000 rpm for 15 min at 4 °C. The acqueous phase was transferred into a clean tube and the RNA was precipitated with 0.5 ml of isopropanol, shaken vigorously by hand for 10 sec, incubated at R.T. for 10 min and centrifuged at 12,000 rpm for 10 min at 4 °C. The supernatant was carefully removed, the pellet washed with 1 ml of cold 75 % ethanol, and centrifuged at 12,000 rpm for 5 min at 4 °C. The pellet was air-dried for 10 min and resuspended in 20 μl of RNase-free water (Ambion) and 1μl was mixed with 1μl of 5 x RNA loading buffer and 4μl of water, denatured at 65 °C for 5 min and analyzed on 10% formaldehyde - 1% agarose gel.

RNA amplification using SMART7

Total RNA was serially diluted using siliconized microtubes, to assess the reliability of the RNA amplification protocol starting from amount equivalent to few cells: 10^3 cells (10 ng RNA) and 10^2 cells (1 ng RNA) were compared to not amplified sample 2 x 10^6 cells (20 μg RNA).

The quality and purity of RNA was evaluated by electrophoresis on denaturing 10 % formaldehyde 1% agarose gel as before and by Nanodrop UV spectrophotometer.

For the 1st strand cDNA synthesis:

1 μl of total RNA was mixed with 125 ng of the oligodT SMART7-24 (TGAAGCAGTGGTAACAACGCAGAGTAATACGACTCACTATAGGGAGAAGCTTTTTTTT TTTTTTTTTTTTTTTTVN) denatured heating at 65 °C for 5 min and immediately transferred on ice.

A mixture containing 125 ng of the strand-switching oligo RiboSMART (AAGCAGTGGTAACAACGCAGAGTACGCrGrGrG), DTT 1 mM, dNTPs 1 mM, 1 x 1st strand

buffer (Invitrogen), 20 units of RNase Inhibitor Superase-in (Ambion), 100 units of SuperScript II (Invitrogen) was added to the RNA and incubated at 37 °C for 1 hour, 70 °C for 10 min.

All the 1st strand reaction was used for the 2nd strand cDNA synthesis in a 100 μl mixture containing 1 x ExTaq buffer (Takara), 200 μM dNTPs (Takara), 0.4 μM PCR-SMART primer (TGAAGCAGTGGTAACAACGCAG), 5 units of ExTaq (Takara) and amplified by PCR using the following program 5 min 95 °C, 19 X (5 s at 95 °C, 5 s at 65 °C, 6 min at 68 °C), doing hot start.

The cDNA was purified on a Qiaquick PCR purification column (Qiagen) according to the manufacturer`s instructions and eluted in 30 μl of water. An aliquot of 3 μl was analyzed on 1.5 % agarose gel stained by ethidium bromide. The rest of the eluate was concentrated in speedvac up to 10 μl.

The 1st *in vitro* transcription (IVT) was performed in 20 μl adding to 10 μl of cDNA, 4 μl of 5 x T7 transcription buffer (Promega), 2 μl of NTPs (100 mM, Promega), 0.5 μl of 100 mM DTT, 0.5 μl of RNase OUT (40u/μl, Invitrogen) and 160 units of T7 RNA Polymerase HC (Promega) and incubated at 37 °C for 5 hours.

The 1st amplified RNA (aRNA) was cleaned up using RNeasy Mini kit column (Qiagen) and eluted in 30 μl of RNase-free water (Ambion); an aliquote of 3 μl was analyzed on denaturing formaldehyde 1 % agarose gel as before, then the eluate was concentrated to 10 μl in speedvac. The 1st aRNA with 100ng of random hexamer (Invitrogen) was heated at 70 °C for 3 min and immediately transferred on ice for 2min. A mixture of 2 μl dNTPs (10 mM, Invitrogen), 4 μl of 5 x 1st strand buffer (Invitrogen), 0.5 μl of RNase OUT (40 U/μl, Invitrogen), 2 μl of 100 mM DTT (Invitrogen), 1,5 μl of SuperScriptII (200 U/μl, Invitrogen) was added in a total volume of 20 μl and incubated at R.T. for 2 min and at 37 °C for 3 hours.

At this point, in the case of evaluation of the amplification process by gene-specific RT-PCR, 1 μl of this 1st strand cDNA was amplified by PCR in 100 μl reaction using 0.2 μM of the gene-specific primers GAPDH forw (), GAPDH rev () or alternatively Bgn forw (), Bgn rev (), in a mixture containing 1 x ExTaq buffer (Takara), 200 μM dNTPs (Takara), 5 units of ExTaq (Takara) and amplified by PCR using the following program 5 min 95 °C, 19 x (30 s at 95 °C, 30 s at 60 °C, 1 min at 72 °C), 10 min at 72 °C, doing hot start.

A 2nd strand synthesis was accomplished in 100 μl using all the previous 1st strand reaction and a mixture of 1 x ExTaq buffer (Takara), 200 μM dNTPs (Takara), 1 μg of 1μg/μl T7T24 primer (AGTAATACGACTCACTATAGGGAGAAGCTTTTT TTTTTTTTTTTTTTTTTTVN), 0,4 μl of RNase H (2U/μl, Invitrogen), 10 units of ExTaq (Takara) and incubating in a MWG

thermocycler using the following program 37 °C 10min, 1 min 95 °C, 2 x (5 s at 95 °C, 2 min at 42 °C, 30 min at 68 °C).

The cDNA was cleaned up as before and an aliquot of 3 μl analyzed on 1.5 % agarose gel stained by ethidium bromide.

A second round of IVT was performed as previously described. The 2nd aRNA was also purified and analyzed the same way as before.

The rest of the eluate was concentrated up to 10 μl and stored at -80 °C, then used for the preparation of the microarray probe.

cDNA microarray: probe preparation, hybridizations and scanning

The probe for microarray was prepared as follows: 10μl RNA sample was mixed with

5μg of oligo d(T)$_{20}$ (Anchored oligo d(T)$_{20}$ from Invitrogen, 2.5μg/μl) and 9μg of random hexamer (1.5μg/μl, Invitrogen) and incubated at 65ºC 5min and then put on ice for at least 1min. A mixture of 1μl of (dNTPs + AminoAllyl-dUTP), 6μl of 5x 1st strand buffer (Invitrogen), 3μl of 100mM DTT (Invitrogen), 0.2μl RNase Out (40 U/μl, Invitrogen) and 400 units of SuperScript II (Invitrogen) was added and incubated at 37ºC for 3 hours. After the reverse transcription was completed the RNA was hydrolized with 10μl 1M NaOH at 65ºC for 10 min and immediately neutralized by adding 10μl 1M HCl. The cDNA was purified on Qiaquick PCR purification column (Qiagen) according to the manufacturer's instructions and eluted with 2 volumes of 30 μl of water, then concentrated in speedvac down to 4.5μl.

4.5μl of 100mM NaHCO$_3$ was added to the samples and incubated at R.T. for 15 min.

The aminoallyl-containing cDNA was labeled as follows: the dyes Cy3 and Cy5 (Amersham Pharmacia) were dissolved each in 2μl of DMSO and were mixed to the samples and kept into the dark, at R.T. for 1 hour. 4.5 μl of 4M hydroxyamine were added to each vial and incubated at R.T. for 15 min. 3.5μl of 100mM sodium acetate pH 5.2 were added and the volume filled to 100μl with water. The labeled cDNA was cleaned up with Qiaquick PCR purification column (Qiagen), eluting with two volumes of 40μl of water, and 2μl were analyzed by Nanodrop UV spectrophotometer to measure purity, concentration and Cy3/Cy5 incorporation (pmol/ μl).

The speedvac was used to decrease the 80 μl volume of the samples to 5 μl. The two differently labeled samples (from wild type and mutant Striatal cells) were combined having a total volume of 10μl. cDNA glass slides (SISSAorf2 array) containing 7000 mouse full length-cDNAs from RIKEN Fantom2 collection, each spotted in triplicate, were pre-hybridized in 0.2 % SSC buffer at 55 ºC for one hour. 10 μl of total labelled cDNA were mixed with blocking reagents mixture

(Salmon sperm, polyA RNA?, COTmouse?, tRNA), formamide, and and incubated at 80 ºC for 10 min and the volume was increased up to 80 μl. Then each slide was allocated in the hybridization hermetic chamber, and the solution was poured to cover all the slide surface. Hybridization was performed at 65 ºC for 16 hours in a water bath. The slides were washed in 50 ml Falcon tubes filled with 2 x SSC buffer, 0.1 % SDS (Ambion), 1 mM DTT (Invitrogen), incubated at 55 ºC for 10 min, then the washing solution was discarded and 50 ml of 0.2 x SSC buffer were poured to the slides and incubated at 55 ºC for 10 min. A third washing step was performed at R.T. for 30 min in 50 ml of 0.02 x SSC buffer. The slides were then centrifuged at 2,000 rpm for 5 min to dry and each slide was scanned at 532 nm and 635 nm using GenePix 4100A scanner (Axon) applying different amplification values for the signal in each channel to normalize for the differences in the global intensity of emission. All the slides were kept in the dark and scanned in a short time to avoid the signal degradation due to the quenching.

Statistical analysis of microarray data

The pre-processing (reading of the slide, intra-array normalization and inter-array normalization) of the data was executed on every group independently. The loading, normalization and statistical analysis were performed by using the Gene Pix 5.0.1 software package. The normalization intra array has been performed by using the function "normalizeWithinArrays" applying the LOWESS algorithm: "normalizeWithinArrays(RG,method="loess",bc.method="normexp",offset=50)".
The inter array normalization by using the function "normalizeBetweenArrays" by the application of quantile method: "normalizeBetweenArrays (MA,method="quantile")".
All the statistical analyses have been performed by using the eBayes function from the Gene Pix package. The filters used are the widely accepted: fold change $\leq \log2 (-1)$ or fold change $\geq \log2 (1)$ (which is a fold change of ± 2 on a linear scale) and corrected p-value ≤ 0.05.

Real Time-PCR of differentially expressed genes in striatal cells

To analyze the relative levels of expression of abundant and rare transcripts, Biglican (Bgn) and β-actin genes were analyzed by Real-Time PCR in wild type versus mutant Striatal cells. Single strand cDNA was obtained from 1 μg of purified RNA using the iSCRIPT™ cDNA Synthesis Kit (Bio-Rad, Laboratories, Hercules, CA, USA) according to manufacturer's instructions. RT was performed in a thermal cycler at 25 °C for 5 min, 42 °C for 30 min, 85 °C for 5 min. cDNA was stored at - 20 °C. Real Time quantitative PCR was performed with an iCycler IQ (Bio-Rad Laboratories, Hercules, CA, USA); β-actin was used as an endogenous control to normalize the

expression level of target genes. The primers were chosen using the software Beacon Designer 2.0 (PREMIER Biosoft International, Palo Alto, CA, USA). Primers used were: mouse-β-actin 5'-CCTTCTTGGGTATGGAATCCTGTG –3' (sense); mouse-β-actin 5'-CAGCACTGTGTTGGCATAGAGG –3' (antisense); mouse-Bgn 5'-TTCTGGGACTTCACCTTGGATG –3' (sense); mouse-Bgn 5' TGGCAGTGGCA ACCGAAAG –3' (antisense). Twentyfive nanograms of cDNA was amplified by PCR using primers for the gene of interest and β-actin in separate wells with 1x iQ SYBR Green Supermix and 250nM gene specific sense and anti-sense primers in a final volume of 25 μl for each well. The PCR was performed in 96-well plates, each sample was performed in duplicate. The thermal cycler conditions consisted of 4 min at 95 °C and 40 cycles at 95 °C for 20 s, 60 °C for 20 s and 72 °C for 30 s. In order to verify the specificity of the amplification, a melt-curve analysis was performed, immediately after the amplification protocol, under the following conditions: 1 min denaturation at 95 °C, 1 min annealing at 55 °C, and 80 cycles of 0.5 °C increments (10 s each) beginning at 55 °C. Non-specific products of PCR were not found in any case. A standard curve was generated using a "calibrator" cDNA (chosen among the cDNA samples), which was serially diluted and analyzed both for the gene of interest and β-actin. The iCycle iQ Real Time PCR Detection System Software generated the equation describing the plots of the log10 of the starting quantity (micromoles) of 6 dilutions (240, 24, 2.4, 0.24, 0.024 and 0.0024 ng) of the calibrator cDNA versus the corresponding threshold cycle. The results were normalized to β-actin and the initial amount of the template of each sample was determined as relative expression versus one of the samples chosen as reference (in this case the control sample) which is considered the 1x sample. The relative expression of each sample was calculated by the formula $2x^{\Delta\Delta Ct}$. ΔCt is a value obtained, for each sample, by the difference between the mean Ct value of the gene of interest and the mean Ct value of β-actin. ΔΔCt of one sample is the difference between its ΔCt value and Ct value of the sample chosen as reference (User Bulletin 2 of the ABI Prism 7700 Sequence Detection System).

nanoCAGE –Illumina-Solexa

All the nanoCAGE libraries in this study were prepared with the RNA extracted from three different dopaminergic cell populations (A7, A9, A10), harvested by laser capture microdissection from the Th-GFP/21-31 line of transgenic mice (Matsushita *et al.*, 2002. J Neurochem. 82: 295-304), and kindly provided from Christina Vlachouli. For each type of dopaminergic neurons about 2500 cells were used.

5 to 10 ng of total RNA were heat denatured at 65 °C for 10 min in a final volume of 5 μl with 50 pmol of strand-switching oligonucleotide riBOSS

(TAGTCGAACTGAAGGTCTCCAGCAGrGrG), 10 pmol of random-N15 primer (GTACCAGCAGTAGTCGAACTGAAGGTCTCCTCTN15), 10 pmol of oligodT18 (GTACCAGCAGTAGTCGAACTGAAGGTCTCCTCT18) and then transferred quickly on ice/water mix.

Reverse transcription was performed in a volume of 20 µl adding the following components to reach these final concentrations: 1 x 1st strand buffer (Invitrogen), 0.5 mM dNTPs (Takara), 1 mM DTT (Invitrogen), 0.75 M Betaine (WAKO), 0.41 M D-Threalose (Nacalai Tesque), 3.4 % D-Sorbitol (WAKO) and 400 units of SuperScriptII (Invitrogen), and incubated at 12 °C for 10 min, 50 °C for 45 min, 75 °C for 10 min in a MWG thermocycler. The tubes were then immediately transferred on ice/water mix.

For the 2nd strand synthesis a small scale PCR reaction was performed to evaluate the optimal number of cycles, defined as the last cycle before the intensity of the product ceases to increase. 10 µl aliquotes were taken every two cycles and analyzed on 1 % agarose gel.

2 µl of 1st strand cDNA were amplified in a total volume of 100 µl using a mixture containing 1 x ExTaq Buffer (Takara), 250 µM dNTPs (Takara), 100 nM Sup-Forward primer (TAGTCGAACTGAAGGTCTCCAGC), 100 nM Sup Reverse primer (GTACCAGCAGTAGTCGAACTGAAGGTCTCCTCT), and 5 units of ExTaq (Takara) with the following PCR program 5 min 95 °C, n x (10 s at 95 °C, 15 s at 65 °C, 6 min at 68 °C) and using an hot start. In the case of A9, A10 libraries 28 cycles PCR were performed.

A large scale PCR preparation using all the 1st cDNA (18 µl) to avoid bottlenecks, was performed in 9 reactions of 100 µl. PCR products were cleaned using Qiaquick PCR purification columns (Qiagen) and all of them were pooled together.

Half of the pooled cDNA was digested in 4 reactions of 300 µl each, using 150 units of EcoP15I (NEB), 1 x buffer 3 (NEB), 1 mM ATP (NEB), 1 x BSA (NEB) incubated at 37 °C for 4 hours. The low molecular weight cleavage products were purified through the Microcon YM-100 membranes (Millipore) and the flow-through was concentrated on Microcon YM-10 (Millipore) according to the manufacturer`s instructions.

Equimolar amounts of the oligonucleotide linker up were annealed with the correspondent oligonucleotide linker down, in a water bath from 95 °C cooling down to R.T. to form the Illumina-Solexa adaptors.

Each pair contains a different tissue tag sequence, AG, TC or GT respectively.

AG linker up (NNAGCTGTAGAACTCTGAACCTGT), AG linker down (ACAGGTTCAGAGTTCTACAGCT), TC linker up (NNTCCTGTAGAACTCTGA- ACCTGT), TC linker down (ACAGGTTCAGAGTTCTACAGGA), GT linker up

(NNGTCTGTAGAACTCTGAACCTGT), GT linker down (ACAGGTTCAGAGTT-CTACAGAC).

The Illumina-Solexa adaptors were ligated to 15μl of the EcoP15I cleavage products using 2 μM adaptor and 1200 units of T4 DNA ligase (NEB) in 30 μl and incubated for 16 hours at 16 °C in a water bath.

Optimal number of cycles for the ligation product to be amplified was determined by PCR with 5 μM long P5 primer (AATGATACGGCGACCACCGACAGGTTCAGA-GTTCTACAG), 5 μM long P7 primer (CAAGCAGAAGACGGCATACGATAGT-CGAACTGAAGGTCTCCAG), 1 x ExTaq buffer (Takara), 200 μM dNTPs (Takara), 5 units of ExTaq (Takara). The program was 5 min 95°C, n x (20 s at 95 °C, 20 s at 57 °C, 20 s at 68 °C). 6 PCR reactions in 100 μl for each sample were performed. The excess of the primers was digested with 20 units of Exonuclease I (Takara) at 37 °C for 30 min and then the enzyme was heat inactivated at 55 °C for 15 min. Then the PCR products were purified by electrophoresis on 10 % polyacrylamide gel and the band corresponding to the expected size (112bp) was cut, passed through a syringe to break the structure of polyacrylamide and the DNA was extracted at R.T. in microtube on rotation with 800 μl of 1 x TE buffer over night.

The tubes were centrifuged at 13,000 rpm for 10min and the supernatant was collected. 600 μl of 1 x TE buffer were added to the polyacrylamide in the microtube and let to rotate at R.T. for 1 hour. This step was repeated once more.

Then for each sample, all the collected fractions were pooled together and passed through a Microspin filter (GE Healthcare) to eliminate residual traces of polyacrylamide.

A total of 2ml of filtrate was concentrated to 100 μl passing through a Microcon YM-10 column (Millipore), further purified using a Qiaquick PCR purification column (Qiagen) and eluted in 100 μl of EB buffer (10mM Tris.Cl pH 8.5).

The purity and the concentration of the sample was estimated by Nanodrop UV spectrophotometer and Agilent 2100 Bioanalyzer and an aliquote of 10 μl was analyzed by PAGE to check for the correct size of the recovered DNA.

One Solexa sequencing run was performed using all the channels for the same sample, to normalize against the bias due to channel`s variability.

nanoCAGE –454

All the initial steps of 1st strand synthesis by reverse transcription, 2nd strand cDNA synthesis by PCR amplification and EcoP15I digestion are the same like for nanoCAGE-Solexa.

454-adaptor forward (NNAGCATGAGACCTGTGAGTAG) and 454-adaptor reverse (Bio-CTACTCACAGGTCTCATGCT) were annealed as for Solexa adaptors, to form the 454 adaptor.

15μl of low molecular weight fraction of the EcoP15I digestion product was ligated using 2 μM of 454 adaptor and 1200 units of T4 DNA ligase (NEB) in 30 μl and incubated for 16 hours at 16 °C in a water bath, then the volume of the reaction was increased up to 100 μl with water.

Optimal number of cycles for 1 μl of the ligation product to be amplified was determined by PCR with 5 μM 5`nested PCR primer (Bio-CGAACTGAAGGTCTCCAGCA), 5 μM 454-adaptor reverse primer (Bio-CTACTCACAGGTCTCATGCT), 1 x ExTaq buffer (Takara), 200 μM dNTPs (Takara), 5 units ExTaq (Takara). The program was 5 min 95 °C; n x (20 s at 95 °C, 20 s at 57 °C, 20 s at 68 °C). Then a large scale preparation of PCR product was done in 96 wells plate to ensure enough material for Eco31I digestion. The 96 PCR reactions were pooled and split in 16 tubes and were extracted by using 600 μl of 25:24:1 Phenol-Chloroform-Isoamyl alcohol (Sigma-Aldrich), 600 μl of Chloroform and then precipitated with 2 voll. of cold ethanol, 1/10 vol. of sodium acetate (pH 5.2) and 10 μg of glycogen (Roche). The pellets were washed with 400 μl of cold 70 % ethanol, dry at R.T. and resuspended in 10 μl of 1 x TE buffer each. Then the content of 16 tubes were pooled together.

All the PCR product was digested with Eco31I (Fermentas) in 10 reactions of 300 μl each containing 15 μl of DNA, 1 x buffer G (Fermentas), 150 units of Eco31I, and incubated at 37 °C for 6 hours. After the digestion was completed, 150 μl of streptavidin-agarose beads (Upstate) were added to each reaction and were incubated on rotation at R.T. for 1 hour to sequester the biotinylated digested end-products, leaving in solution the fraction corresponding to the 5` prime tags (32bp).

The tubes were centrifuged at 13,000 rpm for 10min and the supernatants were recovered. Other 300 μl of 1 x TE buffer were added to each tube to wash the streptavidin beads and incubated at R.T. on rotation. After centrifugation the supernatants were recovered as before.

The 5` tags in the supernatants were extracted by 1 vol of (25:24:1) Phenol-Chloroform-Isoamyl alcohol (Sigma-Aldrich), 1 vol. of Chloroform (WAKO), precipitated with 2 voll. of cold ethanol 1/10 vol. sodium acetate 3 M pH 5.2, 10μg glycogen (Roche), and washed with 1 vol. of cold 70 % ethanol, then dried at R.T. and resuspended in a total of 100 μl 1 x TE buffer. Then this was purified by electrophoresis on a 8 % polyacrylamide gel and the band corresponding to the expected size (32bp) was cut, and purified in the same way as before for the Illumina-Solexa tags. In this case the DNA was extracted using 1 x SAGE buffer (LoTE* : 7.5 M ammonium acetate, 125 vol. : 25 vol.; *LoTE buffer 3 mM Tris HCl pH 7.5, 0.2 mM EDTA pH 8) instead of 1 x TE buffer.

A total of 2 ml of filtrate was concentrated to approximately 10 μl in a Microcon YM-10 column (Millipore) according to the manufacturer`s instructions.

The oligonucleotides 454-A short (CCATCTCATCCCTGCGTGTCCCATCTGTTC-CCTCCCTGTCTCAG), 454-A long (P-TGCTCTGAGACAGGGAGGGAACAGA-TGGGACACGCAGGGATGAGATGG), 454-B short (Bio-TEG-CCTATCCCCTG-TGTGCCTTGCCTATCCCCTGTTGCGTGTCTCAG), 454-B long (P-AGCACTG-AGACACGCAACAGGGGATAGGCAAGGCACACAGGGGATAGG) (Operon) were purified by electrophoresis on a denaturing 10% polyacrylamyde-urea gel.   The gels were stained by ethidium bromide, quickly inspected at 302nm UV light and for each oligonucleotide, the band of the expected size was cut and the DNA was extracted as before for the 5`tags using the SAGE buffer. Then for each oligonucleotide all the collected fractions were pooled together and passed through a Microspin filter (GE Healthcare) to eliminate residual polyacrylamide.

A total of 2 ml of filtrate was precipitated with 2 vol. of ethanol, 1/10 vol. of sodium acetate 3 M pH 5,2 and washed with 1vol. 70 % ethanol, dried and resuspended in 50 μl of 1 x TE buffer.

Equimolar amounts of the purified primer pairs (454-A short / 454-A long) and (454-B short / 454-B long) were annealed in 1 x T4 DNA ligase buffer in a water bath from 95 °C cooling down to R.T. to form 454-Terminator A and 454-Terminators B respectively.   An aliquote of 2 μl was analyzed by PAGE.

After quantity evaluation by Nanodrop UV spectrophotometer, the filtrate corresponding to the 5`tags (10 μl) was concatenated in a 20 μl vol. ligation reaction using a ratio 5`tags: 454-terminators (20:1) and a mixture containing 9 μl of Eco31I digested 5`tags, 20 ng 454-Terminator A, 20 ng 454-Terminator B, 1 x T4 DNA ligase buffer (NEB), 5 % PEG 6000 (Fermentas), 2000 units of T4 DNA ligase high conc. (NEB), incubated at 16 °C for 16 hours in a water bath.

A mixture of 170 μl high-salt CTAB solution [1 % CTAB (Sigma-Aldrich), 4 M urea (WAKO), 0.848 M sodium chloride], 2 μl EDTA 0.5 M pH 8 and 30 μl of water was added to the ligation reaction and incubated at 65 °C for 4min, then at R.T. for 10 min and loaded on a GFX PCR DNA purification column (GE Healthcare).

After spinning in a microcentriguge at 13,000 rpm for 1 min the flow-through was discarded and 600 μl of washing buffer (100 mM Tris-HCl pH 7.5, 5 mM EDTA, 300 mM sodium chloride, 60 % ethanol) was poured to the column and centrifuged at 13,000 rpm for 1 min and the flow-through was discarded.

600 μl of 80 % ethanol was poured to the GFX column and centrifuged at 13,000 rpm for 1 min and then the column was transferred to a new tube and the DNA was eluted with 53 μl of pre-heated 65 °C water. The purity and the concentration of the sample was estimated by Nanodrop UV

spectrophotometer and Agilent 2100 Bioanalyzer and an aliquote of 10 μl was analyzed by PAGE and ethidium bromide staining to check for the expected size distribution of the recovered DNA. One 454 sequencing run was performed using the big PicoTiterPlate of the Genome Sequencer 20 system for the same sample.

Bioinformatic Analysis of nanoCAGE data

Briefly, raw sequence data obtained from a full run (8 channels) of Illumina-Solexa sequencer was processed for "tag extraction" to discard unspecific sequences that did not contain the restriction sites used for tag production and the 2 nt bar codes inserted in the oligonucleotides as well as to distinguish among different samples in the mixed library. During tag extraction also the first G were removed as they came from the riBOSS strand-switching primer. A statistical evaluation of the frequency of occurence of each of the 4 bases at each position of the tags has been done together with the construction of an error estimation model for better implementing the extraction process and lowering the fraction of the false positives.

The extracted tags were mapped to the ribosomal DNA coding for rRNAs and the tags that matched those loci were discarded from the rest of the analysis. All the other tags were mapped and aligned to the mouse genome (mm9 assembly) using Kalign software and chosing to have no mismatch in the alignment in a first run, and to tolerate one mismatch in a second run. Then the tags (TSSs) were clustered. Two strategies have been used to cluster TSSs: FANTOM3 style "proximity" tag clustering (TC) and Martin Frith Parametric clustering (PC).

In the "proximity" tag Clustering (TC), TSSs are clustered together if they are separated by 20 or less base pairs regardless of the RNA library they are derived as in Carninci, Sandelin et al. Nat Genet. 2006.

This clustering approach generated a total of 2,068,275 clusters(xxx clusters with a tpm_density over 1).

With the Parametric Clustering (PC), TSSs are clustered using the method described by Martin Frith in "A code for transcription initiation in mammalian genome" (Frith et al 2007). As noted by Martin, a limitation of the proximity tag clustering approach is that TSS can only belong to one cluster while it is evident that one can observe the existence of clusters within clusters.

His approach enable the definition of such clusters hierarchy. In short, the algorithm employed compute the local density (number of tags / length of the segment) of all genomic segments. The ratio of the minimum density value below which the cluster merges with another cluster and the maximum density value, above which the clusters splits into smaller clusters is used to score the

cluster "stability" (and relevance; if the ratio is small then the cluster is not strongly present in the data and could easily vanish with some changes in the underlying tag counts, while if the ratio is big, the cluster is a prominent feature of the data).

This clustering approach generated a total of 4,736,538 clusters (526,461 clusters with a tpm_density over 1). It is important to note that those clusters are overlapping and that it generate clusters within clusters.

A stability bigger or equal to 2 has been chosen in Fritch et al. as a threshold for stable relevant clusters. (Yet all clusters regardless of their stability score are being reported here). Clusters-transcripts associations were estimated according to the following criteria: for each cluster the extent of the overlap with the transcript sets where extracted, as well as the overlap with region spanning [-500bp -1bp] from the start site of each transcript refered to as "proximal promoter". Clusters that do not overlap with any transcript or their proximal promoter region were annotated as "intergenic". For association with RefSeq transcripts, 20,649 RefSeq genes and their genomic coordinates were downloaded from UCSC mm9 release. 436 miRNA annotated in miRBase and their genomic location were retrieve from UCSC. Non coding transcripts sequences were collected from the FANTOM3 ftp site and blatted on the mouse genome mm9. Those with an alignment of less than 15% were discarded. When a single sequence aligned in multiple places, the alignment having the highest base identity was identified. Only alignments having a base identity level within 0.1% of the best and at least 96% base identity with the genomic sequence were kept.

All the information detailed above has been gathered in 2 tab text files (one for each PC and TC clustering methods), together with expression level for each of the A9 and A10 nanoCAGE libraries. A multimapping rescue strategy similar to that used in Faulkner GJ et al 2008, was used to assign a value also to tags mapping in multiple genomic positions. To limit the subsequent analysis to potentially relevant clusters only tag clusters fulfilling the following criteria have being used: tpm_density (cluster tpm / cluster length) bigger than 1 for at least one library. stability_score (for PC clusters) bigger or equal to 2 and ambiguity_score smaller than 1 (thus retaining clusters for which at least one tag isnt a multi mapper).

An extensive manual evaluation and visual inspection of the genomic distribution of the tag clusters specifically associated to A9-A10 differences have been performed to establish case by case the specific relevance of the findings.

RT-PCR Validation of target genes from total Midbrain

Total midbrain was dissected from adult C57Black/6 female mice. The tissue was homogenized in TRIzol reagent (Invitrogen, Carlsbad, CA, USA), and RNA was extracted according to the

manufacturer's instructions. After DNase I treatment (Ambion, Austin, TX) at 37°C for 1 h, RNA was purified on RNA easy mini kit columns (Qiagen, Hilden, Germany), and RNA quality was tested on agarose gel. RT-PCR reactions were performed using the SuperScriptIII OneStep RT-PCR Kit with Platinum Taq polymerase (Invitrogen, Carlsbad, CA, USA).

For each primer pairs, 250 ng of total RNA were used in a total volume of 50 μl according to the manufacturer's instructions and RT-PCR reaction was performed using the following cycling conditions: 30 min at 57 °C, 5 min at 75 °C, 2 min at 94 °C; then 15 sec at 94 °C, 30 sec at 65 °C, 1,5 min at 68 °C for 45 cycles.

For each gene two isoforms corresponding to the mRNAs transcribed from a distal and a proximal promoter were screened using primer pairs specifically designed to discriminate the corresponding cDNAs.

Primer design was done using Primer3 software (v. 0.4.0) (http://fokker.wi.mit.edu/primer3/input.htm) using the following criteria: primer length ranging from 23 to 28 nt, annealing Tm in the interval 60-72 °C, and GC% content comprised among 50-70 %.

For gene, primer pairs were: Snca-distal fwd 5'-CTTggCACTCAAATCCACTCTgC-3' and Snca-distal rev 5'-CCAgCgAACAgACTCTTCTTCCA -3' [for mouse Snca, RefSeq ID: NM_009221]; Snca-proximal fwd 5'-CCTCCCTAggAAgAggAgCgAAg -3' and Snca-proximal rev 5'-CACACggCTCCCTAggCTTCTgA -3' [for Snca, RefSeq ID: NM_001042451]; Slc18a2-distal-fwd 5'- gAgATATAACCCTgCAggCAgTCg-3' and Slc18a2-distal-rev 5'-CTgTACAggTAgCTgggAATgATgg-3' [for mouse Slc18a2/Vmat2, RefSeq ID: NM_172523]; Slc18a2-intronic-fwd 5' TCTggTTCAgggTCTgTgCACTAAg 3' with either Slc18a2-intronic-rev2 5` AgCCTTTgACTCCTgAgCTCTgTT 3` or Slc18a2-intronic-rev3 5` gggTTTTCCTTgggAgAAAgTCTg 3` [for a transcript corresponding to a new unannotated isoform of mouse Slc18a2, NM_172523 starting in the third intron at position chr19: 59,339,321 in mm9 assembly]; Comt-distal-fwd 5` gACACTCgCACAggCACCCTCTC 3` and Comt-distal-rev 5` AACAggAgACCCAATgAgACAgCAg 3` [for mouse catechol-Omethyltransferase Comt long isoform, RefSeq ID: NM_007744]; Comt-proximal-fwd 5` ggTATATAAAggCTCAggCCCAgTg 3` and Comt-proximal-rev 5` AgAgTCTTTAgAggAggCAggTCCA 3` [for mouse catechol-O-methyltransferase Comt short isoform, RefSeq ID: NM_001111063]; Slc6a3-distal-fwd 5` CgATAAgAgCTCAAggCTgAgACC 3` and Slc6a3-distal-rev 5` TTCTgCTCCTTgACCAAgATgAgC 3` [specific for distal 5`UTR of mouse Slc6a3; RefSeq ID: NM_010020]; Slc6a3-proximal-fwd 5` CCTCATCTCTTTCTACgTgggCTTC 3` and Slc6a3-proximal-rev 5` CAgCACACCACgCTCAAAATACTC 3` [for the internal part starting from exon

4 of mouse Slc6a3 NM_010020]; S-Ubc-fwd 5` AgCCCAgTgTTACCACCAAg 3` and S-Ubc-rev 5` gCAAgAACTTTATTCAAAgTgCAA 3` [for mouse polyubiquitin/Ubc RefSeq ID: NM_019639]; AS-Ubc-rev 5` gACAggCAAgACCATCACC 3` and S-Ubc-rev as before [for an antisense transcript overlapping the Ubc locus and having multiple starting sites].

For mitochondrial fission/fusion genes the following primer pairs were used: Drp1 1stfwd 5` CTgATCCCggTCATCAATAAgCTg 3` and Drp1-2ndrev 5` ACgTgg ACTAgCTgCAgAATgAgA g 3` [for mouse Drp1 RefSeq ID: NM_134850]; Fis1 1stfwd 5` gAgACTgTggCCCAgTAgAgACCTT 3` and Fis 2ndrev 5` CCTCTgC ggATgTCCTCATTgTAT 3` [for mouse Fis1, RefSeq ID: NM_025562]; Opa1 9- 10thfwd 5`-gCTCTCCAgTgAAggTgACTC TCAg-3` and Opa1 11-12threv 5` AgAgATATggTCTCggggCTAACAg 3` [for mouse Opa1, RefSeq ID: NM_133752]; March5-1stfwd 5`- AAgCCCTTCAACAgATgCTggAC-3` and March5-2ndrev 5` TTAAgTACTCggCgTTgCACTgAg-3` [for mouse March5, RefSeq ID: NM_027314]; Sh3glb2-1stfwd 5` ggACgCgggCATCTTCTTCACTC 3` and Sh3glb2-3rdrev 5` gCACCTTTCTgTCCAgCTTCTCAT 3` [for mouse Sh3glb2, RefSeq ID: NM_139302].

Total RNA from striatum used for negative controls was purified as before specified from the same C57Black/6 female mice.

5`-RACE validation of target transcripts starting at specific TSSs

The total midbrain was dissected from adult C57Black/6 female mice and the RNA was extracted and treated with DNAse as described before for RT-PCR analysis. 5μg of total RNA were used for each 5`RACE using the GeneRacer kit (Invitrogen, Carlsbad, CA, USA) according to the the manufacturer's instructions.

100ng of random primers were used for the reverse transcription step. 1 μl of the RT reaction was used for the first 5`RACE-PCR, and Platinum Taq -High Fidelity DNA Polymerase (Invitrogen, Carlsbad, CA, USA) was employed, performing touchdown PCR with the following cycling conditions: 2 min at 94 °C, then 30 sec at 94 °C, 2 min at 68 °C for 5 clycles; then 15 sec at 94 °C, 30 sec at 65 °C, 2 min at 68 °C for additional 25 cycles. The gene specific primers used are listed herein below.

The resulting PCR products were checked on agarose gel and the samples which did not give rise any amplified visible products, were subjected to a second round of nested PCR (25 additional cycles) following the same conditions used for the first round PCR. Then the PCR reactions were separated on agarose gel and purified using the Qiaquick gel extraction kit (Qiagen, Hilden,

Germany), and directly cloned in either the pCR4-TOPO cloning vector (Invitrogen, Carlsbad, CA, USA) or in pGEM-T easy vector (Promega, Madison, WI, USA). DH5-α chemical competent bacterial cells were transformed and 5-10 colonies per plate were screened for the presence of an insert of the correct size by EcoRI (NEB, Ipswich, MA) digestion of the corresponding plasmid DNA purified on Qiagen Miniprep columns (Qiagen, Hilden, Germany). Each positive clone was sequenced by the Sanger method with the M13 rev primer using the SequiTherm EXCEL™ II DNA Sequencing Kit (Epicentre Biotechnologies, Madison, WI, USA). The sequences were then aligned to mouse genome (mm9 release) using the BLAT implementation on the UCSC Genome Browser and scored by the percentage of identity, their length and their correspondence to the annotated gene structure (splicing pattern).

In Situ Hybridization (ISH) analysis of transcript isoforms on mice midbrain cryosections

For riboprobes preparation plasmid DNA (pCR4-TOPO or pGEM-T easy clones) was digested with an appropriate restriction enzyme to create a linear DNA template for digoxigenin-labelled riboprobe synthesis and then purified on PCR purification coulumns (Qiagen. 1 μg of linearised template DNA was transcribed *in vitro* in 20 μl in presence of 2 μl of 10x DIG-lageling mix (Roche diagnostics GmbH, Mannheim, Germany), 20 units of RNaseOUT inhibitor (Invitrogen, Carlsbad, CA, USA), and 10 units of T3 or T7 RNA polymerase (Ambion, Austin, TX) or alternatively SP6 (Promega, Madison, WI, USA). Transcription was performed for 2 hours at 37ºC. The riboprobes were then purified by addition of 2.5M lithium chloride and 2.5 volumes of ethanol. After centrifugation for 30 minutes at 4ºC, the pellet was washed in ethanol 75% and air dried and resuspended in 50μl nuclease-free water (Ambion, Austin, TX). Riboprobes quality was assessed on agarose gel and their concentration estimated at ND-1000 spectrophotometer (Nanodrop technologies, Wilmington, DE, USA). Adult female C57Black/6 mice were anesthetized with 0.5 ml urethane (100mg/ml) and perfused transcardially with 4% paraformaldehyde (PFA) filtered with a 0,22 μm filter (Millipore). Then the brain was removed and the midbrain region was dissected, rapidly washed in sterile PBS buffered solution (pH 7.4) and fixed for 2-3 hours at 4 °C and then left overnight in a 30% sucrose solution (Sigma-Aldrich, St. Louis, MO, USA) for cryoprotection. The day after, brain were included in OCT medium and rapidly frozen using a mixture of 2-methylbuthane and liquid nitrogen and 18 μm-thick coronal sections were cut on a cryostat. Sections were air-dried for 2 hours, fixed with 4% PFA for 10 min at R.T. and then washed twice in diethyl pyrocarbonate (DEPC, Sigma)-treated PBS for 5 min at R.T. After 5 min treatment with HCl 0.2M, the slides were whased three times in PBS and digested 15 min at 37 °C with 0.5 μg/ml

proteinase K (Roche diagnostics GmbH, Mannheim, Germany) in pK buffer (TrisHCl pH8, EDTA pH8). After washing twice in PBS 5 min, the slides were re-fixed with fresh 4% PFA 10 min at R.T. and then washed again in PBS. Acetylation was done in TEA buffer (Triethanolamine, 27mM HCl) adding twice 400 μl of acetic anhydride for 10 min, then washed with milliQ water and air dried for 1 hour. 500 ng of each ribobrobe were added to 150 μl of hybridization mix (composition for 10ml: 1ml salt mix 10x, 0.5ml DTT 1M, 0.5ml polyadenilic acid -10mg/ml, 50 μl tRNA – 10.7 mg/ml, 2ml dextrane sulphate 50%, 5 ml formamide), heat denatured 10 min at 80 °C and then incubated in a humid chamber (made with 2xSSC-50% formamide) overnight with each slide at 63 °C, keeping separated sense and antisense riboprobes. The day after post-hybridization washes were done with increasing stringency: 30 min et RT in 5x SSC–β-mercaptoethanol, 30 min at 60 °C with 2x SSC- 50% formamide β-mercaptoethanol, twice 15 min at 37 °C in NTE buffer (); finally 15 min at RT in 2x SSC and 15 min at RT in 0.2x SSC buffer. After two additional washing in B1 buffer (), the slides were blocked for 1 hour at RT using heat-inactivated fetal calf serum (FCS) (1:9 in B1 buffer) and then incubated overnight at 4 °C with 150 μl of a mix of blocking solution containing anti-digoxigenin antibody (Roche diagnostics GmbH, Mannheim, Germany) diluted 1: 1000. After washing in B2 buffer, the slides were incubated with levamisol 2 mM in B2 buffer for 1 hour to inhibit the endogenous alkaline phosphatases. A stock solution of Tris-NaCl-Polyvinyl alcohol by dissolving 10% (w/v) polyvinyl alcohol (PVA, 70-100 KDa, Sigma-Aldrich, St. Louis, MO, USA) at 90 °C in 100 mM Tris HCl pH 9 and 100 mM NaCl.

Then the slides were incubated with a BCIP-NBT-PVA development solution containing 0.35 mg/ml NBT (4-nitro blue tetrazolium chloride, Roche) and 0.35 mg/ml BCIP (5-bromo-4-chloro- 3-indolyl-phosphate, Roche). The slides were visually inspected under the microscope and when colour had developed to the desired extent, embryos were washed with B2 solution with 2mM EDTA and slides were prepared for the immunofluorescence analysis with an anti-Tyrosine Hydroxylase (TH) antibody (TH-16 Sigma-Aldrich, St. Louis, MO, USA) as described herein after.

Immunofluorescence (IF)

Adult C57Black/6 mice were perfused transcardially as previously described for ISH. Dissected midbrain was removed and post-fixed in 4% formaldehyde in 0.1 M phosphate buffer pH 7.4, for 2 h at 4°C, equilibrated in 30% sucrose overnight, and 18 μm-thick coronal sections were cut on a cryostat. Sections were air-dried for 30 min or even stored at -80 °C up to 2 months before use, blocked 1 hour at RT with 10% normal goat serum, 1% BSA, 1% fish gelatin in PBS, and incubated overnight at 4°C with the primary antibodies diluted in PBS, 1% BSA, 0.1% fish gelatin, 0.3%

Triton X-100. (For anti-TH antibody dilution 1:1000 was used whereas other primary antibodies were used following the manufacturer`s instructions).

After rinses in 0.1% Triton X-100 in PBS, sections were incubated with fluorophore-conjugated secondary antibodies in the same solution as for primary antibodies for 2 h at room temperature, washed and mounted with Vectashield mounting medium (Vector Laboratories, Burlingame, CA). The staining was analyzed with a Leica DM500 digital microscope equipped with a DFC digital camera. Primary antibodies were: mouse monoclonal anti-TH (Sigma-Aldrich, St. Louis, MO, USA) was used at 1:1000; goat anti-Endophilin B2 (C-16 - SantaCruz Biotechnology Inc., Santa Cruz, CA), goat anti-Fis1 (K-14 - SantaCruz Biotechnology Inc.) and goat anti-Opa1 (C-15 - SantaCruz Biotechnology Inc.) all used at 1:50. Secondary antibodies were: Alexa 488-conjugated donkey anti-mouse, Alexa-594-conjugated donkey anti-goat, and Alexa 594-conjugated goat anti-mouse diluted 1:250 (Molecular Probes-Invitrogen). Nuclei were labelled with DAPI diluted 1: 10,000.

Oligonucleotide sequences and functional grouping

SMART7 primers:
```
TGAAGCAGTGGTAACAACGCAGAGTAATACGACTCACTATAGGGAGAAGCTTTTTTTTTTT
TTTTTTTTTTTTTVN                        SMART7-24
AAGCAGTGGTAACAACGCAGAGTACGCrGrGrG              riboSMART
TGAAGCAGTGGTAACAACGCAG                         PCR-SMART
AGTAATACGACTCACTATAGGGAGAAGCTTTTTTTTTTTTTTTTTTTTTTTTVN
T7T24
```

RT Primers for nanoCAGE
```
          TAGTCGAACTGAAGGTCTCCAGCAGrGrG          riBOSS
GTACCAGCAGTAGTCGAACTGAAGGTCTCCTCTNNNNNNNNNNNNNNN
                                          random N15
GTACCAGCAGTAGTCGAACTGAAGGTCTCCTCTTTTTTTTTTTTTTTTTT
                                          oligodT18
```

Second Strand cDNA synthesis Primers:
```
TAGTCGAACTGAAGGTCTCCAGC                     Sup-Forward
GTACcagcagTAGTCGAACTGAAGGTCTCCTCT           Sup-reverse
```

Solexa Ligation Adaptors and PCR Primers:
```
NNAGCTGTAGAACTCTGAACCTGT      AG linker up    - Solexa
ACAGGTTCAGAGTTCTACAGCT        AG linker down  — Solexa
NNTCCTGTAGAACTCTGAACCTGT      TC linker up    - Solexa
```

```
ACAGGTTCAGAGTTCTACAGGA          TC linker down  — Solexa
NNGTCTGTAGAACTCTGAACCTGT        GT linker up    — Solexa
ACAGGTTCAGAGTTCTACAGAC          GT linker down  — Solexa
AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAG          long P5
CAAGCAGAAGACGGCATACGATAGTCGAACTGAAGGTCTCCAG      long P7
```

454 Ligation Adaptors and PCR Primers:
```
NNAGCATGAGACCTGTGAGTAG                  454-Adaptor fwd
Bio-CTACTCACAGGTCTCATGCT                454-Adaptor rev
Bio-CGAACTGAAGGTCTCCAGCA                SR 5'Nested PCR
```

454 Concatenation Terminators:
```
CCATCTCATCCCTGCGTGTCCCATCTGTTCCCTCCCTGTCTCAG
                                        454-A short
P-TGCTCTGAGACAGGGAGGGAACAGATGGGACACGCAGGGATGAGATGG
                                        454-A long
Bio-TEG-CCTATCCCCTGTGTGCCTTGCCTATCCCCTGTTGCGTGTCTCAG
                                        454-B short
P-AGCACTGAGACACGCAACAGGGGATAGGCAAGGCACACAGGGGATAGG
                                        454-B long
```

5`RACE-PCR and nested-PCR Primers:
```
gTTTCCTTTgTgATCTCCCTgTCC                Slc6a3 3utrrev
ATgTAgCAgCTggAACTCATCgAC                Slc6a3 10threv
ggAgAAgCTCgTCAgggAgTTAAT                Slc6a3 8threv
CACAgAgACggTAgAAgTCCACACT               Slc6a3 6threv
CTgTACAggTAgCTgggAATgATgg               Slc18a2 distal rev
AgCCTTTgACTCCTgAgCTCTgTT                Slc18a2 intronic rev2
gggTTTTCCTTgggAgAAAgTCTg                Slc18a2 intronic rev3
ATCATTCCCAgAAgAgCACACAgC                Slc18a2 11threv
gAACACATggTCTCCATCATCCAg                Slc18a2  10threv
gCgTTACCCCTCTCTTCATCATCTg               Slc18a2 6threv
CCAgCgAACAgACTCTTCTTCCA                 Snca distal rev
CACACggCTCCCTAggCTTCTgA                 Snca proximal rev
AACAggAgACCCAATgAgACAgCAg               Comt distal rev
AgAgTCTTTAgAggAggCAggTCCA               Comt proxi rev
AAgCggAgTTCATAgAgCggAAC                 Jtv1 4threv
CgTACTCTTCACACACAgCgTCACT               Pms2 7threv
gACAggCAAgACCATCACC                     AS Ubc rev
```

164

SECTION 3. RESULTS

cDNA Microarray transcription profiling from limiting amount of starting material

Overview and Validation of the SMART7 technique

A study comparing the classical T7-based method with SMART-PCR on cDNA arrays found that,gene expression measurements of linearly amplified material showed better correlation with non amplified samples than material from SMART-PCR (Puskas et al. 2002) while in a more recent study SMART showed to have the highest true positive discovery rate (80%) compared to a more modest rate reached by a T7-based method (38.6%) (Subkhankulova and Livesey 2006).

This suggests that a method which could take the advantages of both T7-based IVT and SMART-PCR methods could result in a more effective strategy to better amplify small size samples RNA.

SMART (Switching Mechanism At the 5'end of RNA Transcripts) was originally developed as a strategy to enrich for full-length in cDNA cloning. This technique has been used in combination with PCR to amplify minute samples obtained from sources such as biopses, laser captured microdissected tissues, embryos and small organisms (Puskas et al. 2002; Petalidis et al. 2003; Wang et al. 2003; Ji et al. 2004; Kenzelmann et al. 2004; Rox et al. 2004).

Various modification of the SMART-PCR strategy have been reported so far.

The SMART7 method has been previously developed to explore the transcriptome of single, genetically labelled dopaminergic neurons in the mouse retina (Gustincich et al. 2004). It combines an initial cDNA synthesis step based on SMART-PCR with T7 *in vitro* transcription (figure 3-1). The number of PCR cycles as well as the number of rounds of T7 linear amplification can vary according to the starting amount of input RNA.
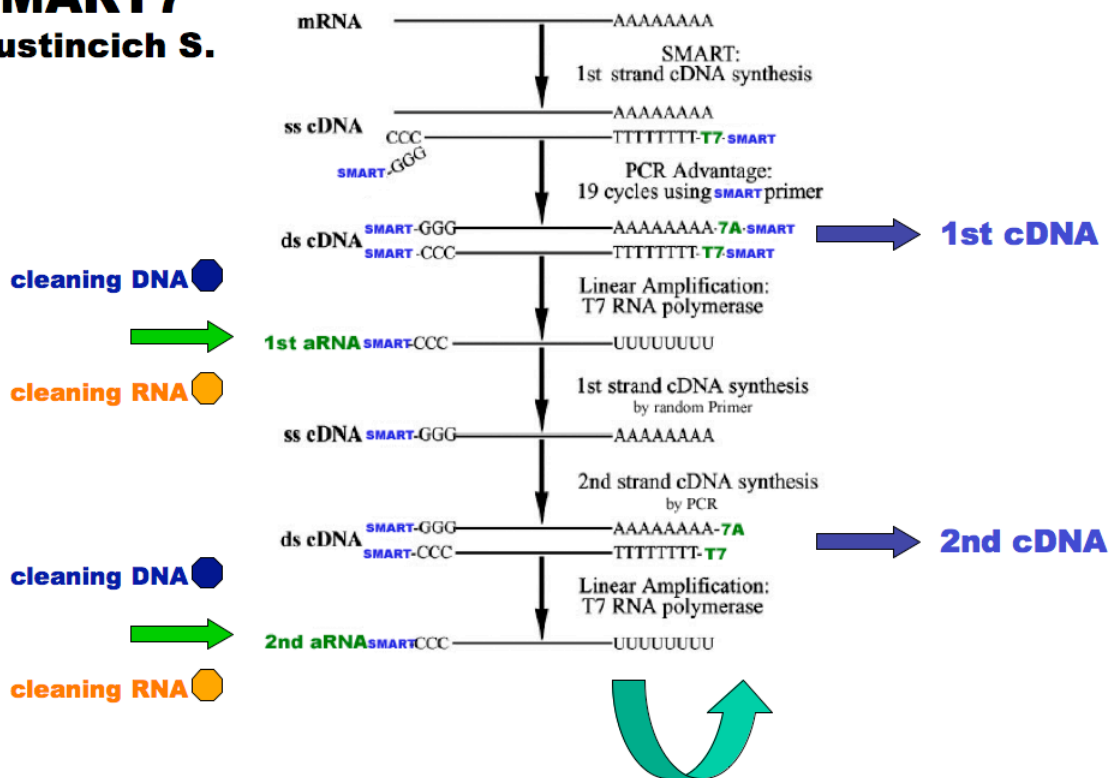
Figure 3-1. Overview of the SMART7 protocol: total RNA is reverse transcribed by mean of a modified oligodT bearing a heel sequence with minimal promoter for T7 RNA polymerase in the presence of a strand-switching SMART oligonucleotide; the 1st strand cDNA is then amplified by single-primer PCR for 19 cycles. The 2nd strand cDNA is purified and subjected to two rounds of *in vitro* transcription to obtain a 2nd cDNA pool. (picture adapted from Gustincich et al 2004).

In previous studies (Luo et al. 1999; Wang et al. 2000; Baugh et al. 2001) it has been shown that two rounds of *in vitro* transcription are required for profiling 1,000 cells or 2ng of total RNA as starting material. However, template-independent artefacts became prominent when smaller amounts of RNA were used and these strongly competed with the specific amplification of the transcripts (Baugh et al. 2001).

In SMART7 the combination of two different techniques kept the number of the PCR cycles low and avoided strong competition from template-independent products. Furthermore, SMART-SMART PCR products, formed by unspecific internal priming of the SMART oligonucleotide during cDNA synthesis, were not amplified by T7 RNA polymerase (Gustincich et al. 2004).

Here I investigated the reproducibility and the fidelity of this technique finding out which is the lower limit of application.

The SMART7 technique was applied with minor variations.

For the first strand cDNA synthesis step, scaling amounts of RNA (100ng, 10ng, 1ng, 100pg) were mixed to 125 ng of the SMART7-24 oligonucleotide (TGAAGCAGTGGTAACAACGCAGAGTAATACGACTCACTATAGGGAGAAGCTTTTTTTT TTTTTTTTTTTTTTTTVN), a modified oligodT primer which contains in the 3` end a stretch of 24 dT followed by the sequence for T7 RNA polymerase promoter and by a SMART tag at the 5` end.

Another oligonucleotide, called riboSMART (AAGCAGTGGTAACAACGCAGAGTACGCrGrGrG), was added to the first strand synthesis mixture. It contained a stretch of three ribo-guanosines that annealed to the dC residues preferentially added by the reverse transcriptase at the 5`end of the first strand cDNA. This occurs in correspondence of the cap structure by a mechanism not yet completely understood, involving a terminal nucleotidil transferase activity.

By means of a strand-switching effect, the reverse transcriptase changed template in correspondence of the annealed riboSMART oligonucleoide, thus incorporating its sequence into the 3`end of the newly synthesized cDNA.

The second strand cDNA was then synthesized by a long cycle PCR using a single primer amplification strategy (PCR-SMART primer, see Materials and Methods for more details) to select for full-length molecules. To minimize the side-effects deriving from exponential overcycling, the number of PCR cycles has been kept as low as possible.

Furthermore, the use of a single primer for the second strand synthesis resulted in a sort of semi-linear amplification behaviour, similarly to what have been seen in other studies [single primer amplification (SPA) as in Smith et al. 2003].

This ensured that all the SMART-tagged full-length cDNAs containing the T7 RNA promoter sequence in the 5`end would be subsequently transcribed, resulting in a 1st round of aRNA.

A second cDNA synthesis step was performed on the 1st *in vitro* transcription product. By using a random hexamer the 1st strand was synthesized. A T7T24 primer that annealed on the poly(A) tail

and on the T7 RNA polymerase promoter sequence was used to synthesize the 2nd strand cDNA. RNAse H was included to degrade the RNA in the mRNA/cDNA hybrids and exTaq DNA polymerase to extend double strand cDNAs. In a first attempt to test the methodology, the gene coding for beta actin cloned in the plasmid vector of the Ambion Maxiscript Kit was used as a template to produce a single *in vitro* transcribed species (pTRI IVT). The transcript was quantified and serial dilutions were used as starting material to perform the SMART7 amplification.

The amplification factor estimated as a ratio between the initial amount of mRNA and the RNA produced by the first and second IVT step of the SMART7 protocol was 10^2 and 10^4 respectively , Starting from 1ng of a single RNA species (pTRI IVT), this is evaluated by ethidium bromide staining of agarose RNA gel electrophoresis and by semi-quantitative PCR evaluation of the cDNA produced after a first and a second reverse transcription steps (figure 3-2).
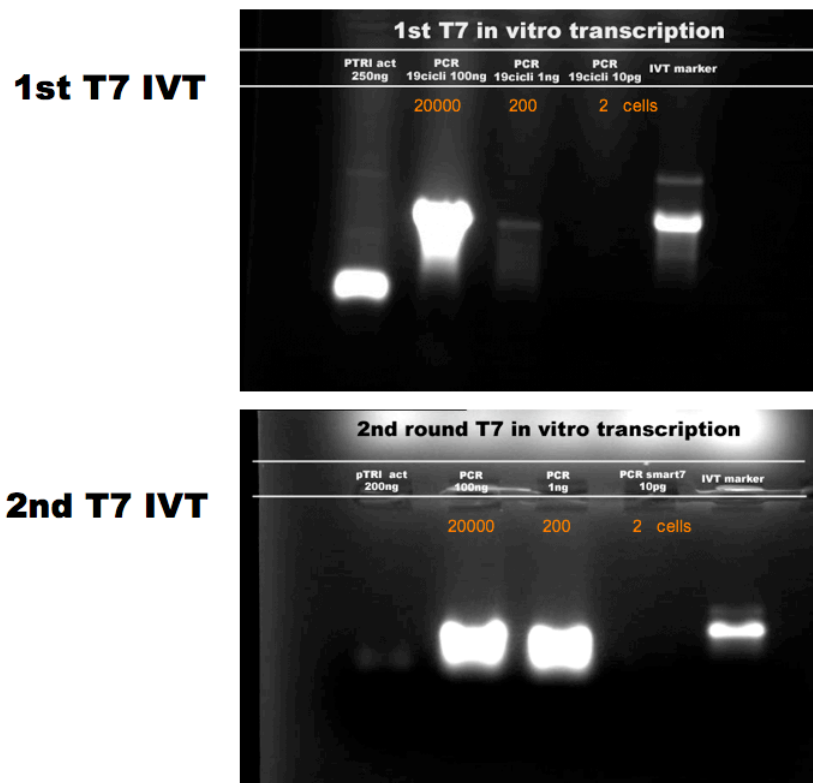
Figure 3-2. SMART7 amplification of a single *in vitro* transcribed RNA species (pTRI act) starting from different initial amonunts (100 ng, 1 ng, 10 pg) as calculated from serial RNA dilutions, and corresponding to the RNA content of approximately 20,000 – 200 -2 cells respectively. A 10^4-fold increase is achieved after 2 rounds of *in vitro* transcription.

After the method has been proved to be effective in the amplification of single mRNA species, SMART7 was applied to the amplification of total cellular RNA, which represents the actual target fot the application of this technique. This method was in fact originally developed to circumvent the limitation derived from the need for a higher amount of RNA in gene expression microarray profiling of single neurons.

Overview and Validation of the "Brownstein method"

As said before, two main different strategies have been applied for sample amplification, one based on linear amplification methods and the other based on exponential amplification methods.

Among the linear amplification studies, we chose to focus our attention on the method introduced by Brownstein (Xiang et al. 2003). They used oligo dT with a T7 RNA polymerase recognition site at the 5'end to prime the first cDNA synthesis step, and a random 9-mer with a T3 RNA polymerase recognition sequence at the 5'end, to prime the two subsequent RT reactions. Three rounds of *in vitro* transcription were carried out (Xiang et al. 2003) (figure 3-3).
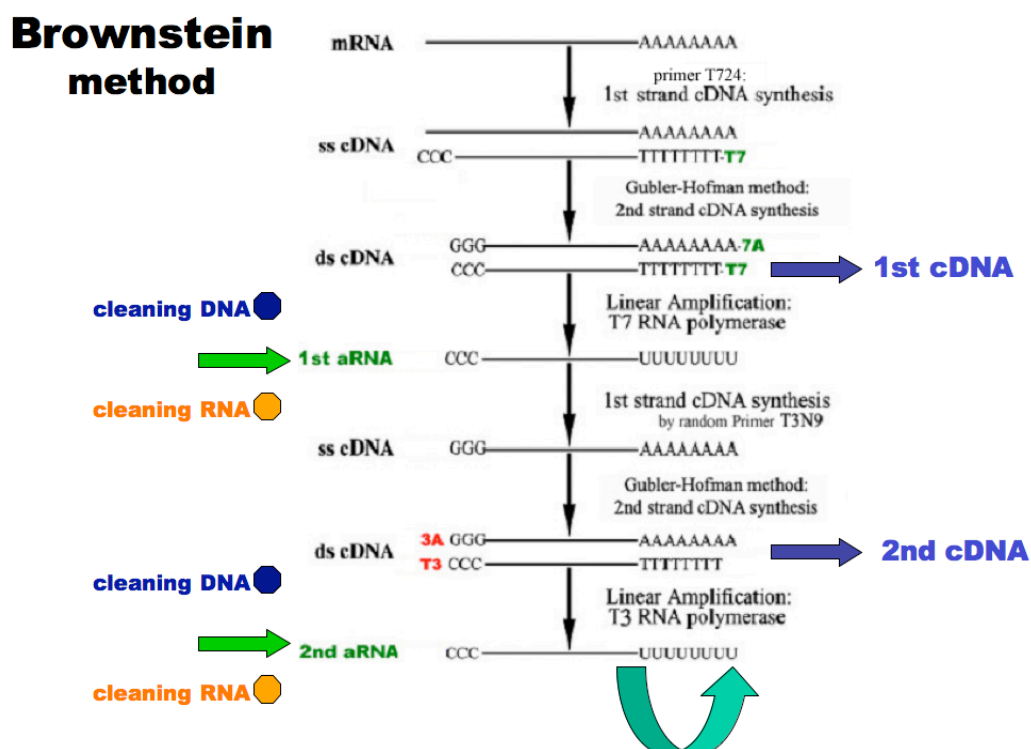
Figure 3-3. Overview of the Brownstein method – The 1st strand cDNA is synthetized in the presence of the classical T7-oligodT primer, the 2nd strand cDNA is produced using the classical Gubler-Hoffman protocol, then follows a 1st round of T7-mediated *in vitro* transcription, a second RT in the presence of a random 9-mer bearing the recognition sequence for T3 RNA polymerase. A second round of T3-mediated *in vitro* transcription and subsequent RT give the 2nd cDNA pool.

In order to evaluate if the SMART7 method performed at a similar level in terms of linearity and of global yield as the transcription-based Brownstein method, cDNA obtained from each reverse transcription step was used as template for gene-specific PCR reactions.

In a first experiment, 100ng of the single *in vitro* transcribed RNA (pTRI IVT) was used to be amplified by using either Brownstein protocol or SMART7, and an aliquote of the resulting 1st strand and 2nd strand cDNA was directly loaded on agarose gel and stained by ethidium bromide (figure 3-4).

Figure 3-4. Comparison between SMART7 and Brownstein amplification methods – A first quantitative evaluation of the yield for the amplified 1st strand and 2nd strand cDNAs resulting from either one of the two methods, starting from the same amount (100 ng) of the single *in vitro* transcribed RNA species (pTRI IVT). Furthermore there si no significant loss of material in the purification step of the ds-cDNA on Qiagen columns (pre-cleaning / post-cleaning).

As shown in figure 3-4, while the first strand cDNA synthesis performed at the same level for both the methods, in the 2nd strand cDNA synthesis step the SMART7 method has been showing to yield an amount of cDNA several orders greater than what obtained with the Brownstein protocol. This is expected due to the PCR exponential amplification, which is not present in the Brownstein method. To address potential alteration of the reciprocal ratio of level of expression, the relative expression of two transcripts already known to be present in striatal cells at very different levels, biglican (Bgn) and glyceraldehyde dehydrogenase (Gapdh), has been evaluated. Quantitative RT-PCR was carried out on aliquots of the cDNA taken after each cDNA synthesis step comparing three different amplification protocols: Brownstein method, SMART7 and a modified version of the SMART7 (S*) in which the 2nd strand synthesis step was performed using the classical Gubler-Hoffman method (a mix of RNase H, E. coli DNA polymeraseI instead of ex Taq DNA polymearase on the random hexamer primed cDNA) (figure 3-5). Although the great difference in the yield for the 2nd strand cDNA due to the exponential amplification in SMART7, no significant difference in the relative transcripts ratio representation was observed since the intensity of the amplicons for all of three compared methods was fairly equal. Furthermore, when we estimated the reproducibility of three different replicas of the amplification starting from the same sample of total striatal RNA, we observed a very high score,, as shown in figure 3-5. These results suggest that the PCR in the SMART7 method, at least for the genes examined, have not introduced any significant bias which could alter the relative expression ratios of the transcripts.

Figure 3-5. Evaluation by gene-specific RT-PCR of the performance of three different amplification methods: Brownstein (B), SMART7 (S) and a modified SMART7 (*) in which 2nd strand cDNA was synthetized by the Gubler-Hoffman method. Mouse biglican (Bgn) was used as low expressed gene, and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a basal transcribed gene. Total RNA from striatal cells was used. As shown in the upward scheme, a 1/20 aliquote of the cDNA after each of the 4 steps (1) pre-incubation before RT, (2) the 1st dscDNA after the first amplification step, (3) the 2rd ds-cDNA after the 1st round of IVT, (4) the 3rd ds-cDNA after the 2nd round of IVT. PCR amplification with gene-specific primers was performed and 5 μl aliquots were taken after either 28 or 40 cycles, and loaded on agarose gel.

Evaluation of total cellular RNA amplification by gene-specific PCR
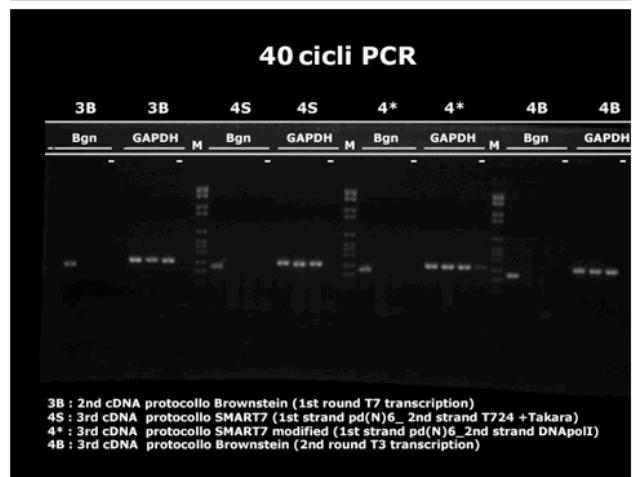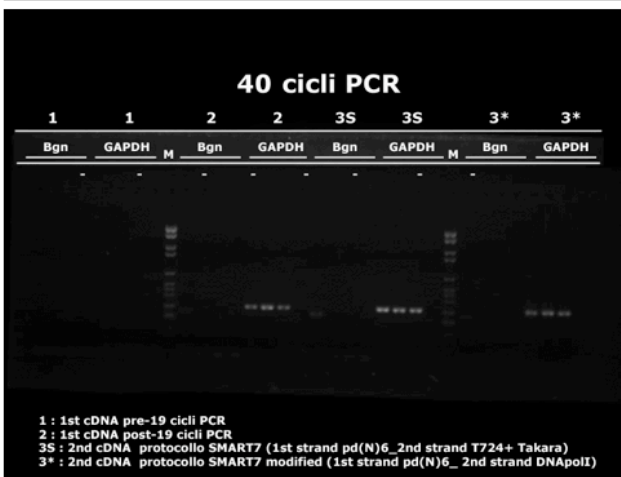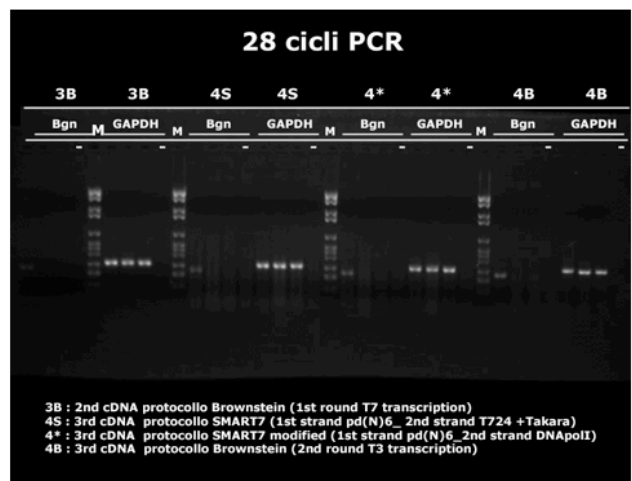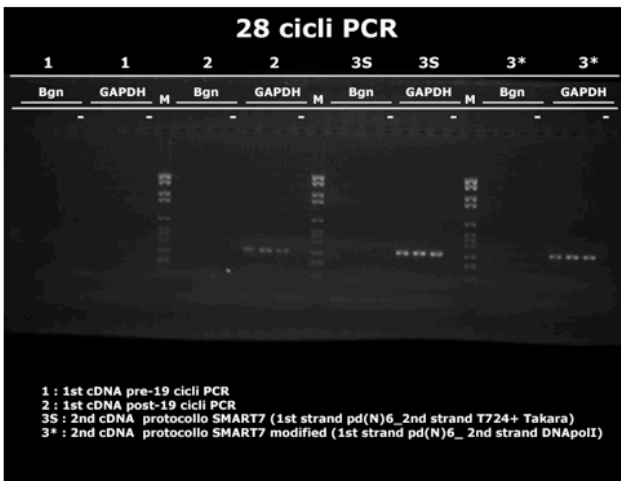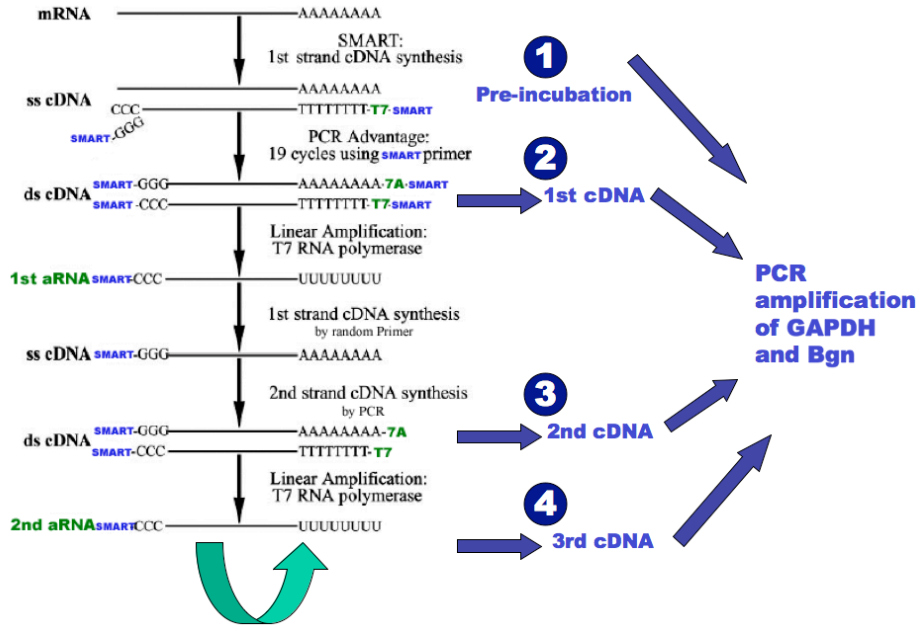
SMART7- and cDNA microarrays

After the preliminary gene-specific RT-PCR-based evaluation of the SMART7 method, a more detailed evaluation on a much larger scale was necessary to assess the validity of SMART7 as a method of choice for the amplification of total RNA obtained from limited number of cells. In order to do this, two lines of cultured cells have been chosen as a source of RNA. Total RNA was extracted from *wild type* striatal neurons and mutant striatal neurons (STHdh^{Q111}/Hdh^{Q111}), from knockin mice bearing the HD mutation (Trettel F et al 2000; kindly provided by Persichetti F). These two cell lines have been previously used in our group for global gene expression profiling using the SISSA orf1 and orf2 glass slides, each containing about 7000 mouse full length-cDNAs from the RIKEN Fantom2 collection, each spotted in triplicate.

Starting from 20μg of high-quality total RNA extracted from *wild type* and mutant striatal cultured neurons, serial dilutions were used to obtain scaling amounts (10ng, 1ng) of RNA used as template for the SMART7 method. Two rounds of *in vitro* transcription have been performed in order to obtain enough aRNA to produce a sufficient amount of probe for microarray hybridization (figure 3-6).

The probes labeled by AminoAllyl-dUTP incorporation and alternatively Cy3 or Cy5 dye, corresponding to *wild type* and mutant striatal RNA, were mixed and hybridized together onto SISSAorf2 cDNA microarray slides. The images were acquired at 532 nm and 635 nm using GenePix 4100A laser scanner (Axon), and the data analyzed with the Gene Pix 5 software package (figure 3-7).

A scatter plot, showing the global correlation in the gene expression profiles corresponding to different starting amounts of imput RNA (10ng, 1ng), was compared to the profile in the case of non amplified sample (20μg) as shown in figure 3-8.
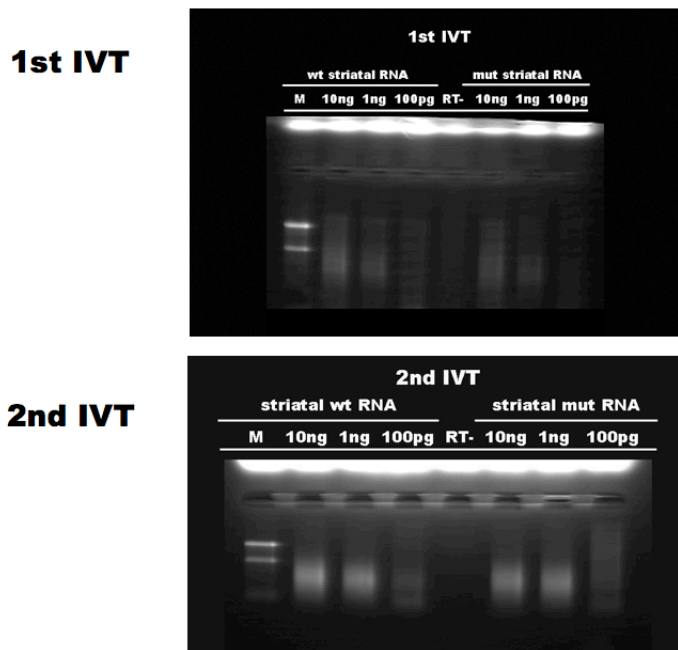
IVT on SMART7 amplified striatal RNA

Figure 3-6. SMART7 amplified aRNA resulting from 19 cycles of PCR and two rounds of *in vitro* transcription, was loaded on agarose RNA gel (1/10 vol. aliquots) and stained with ethidium bromide to compare the yield of 1st and 2nd round amplification, and check for size distribution. The resulting cDNA amplified respectively from 10 ng and 1 ng was labelled for cDNA microarray hybridization. Total RNA from *wild-type* or mutant Striatal cells was used.

Striatal cell's RNA Microarray SISSAorf2
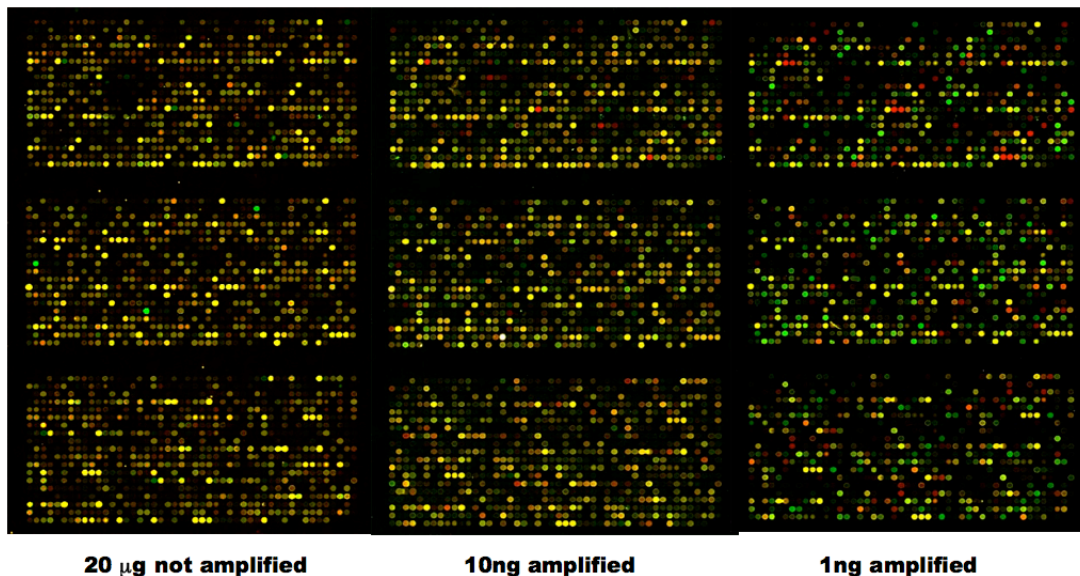
Figure 3-7. Comparison of the fluorescence intensities of spots on SISSAorf2 slide after hydridization and washing steps, as acquired by the GenePix Laser Scanner. An enlargement of a small portion of the slide is represented here, to show the signals for unamplified sample (20 µg) compared to amplified samples (10 ng, 1 ng). Cy5-labelled probe corresponds to *wt* striatal RNA, and Cy3-labelled probe correponds to mutant striatal RNA.
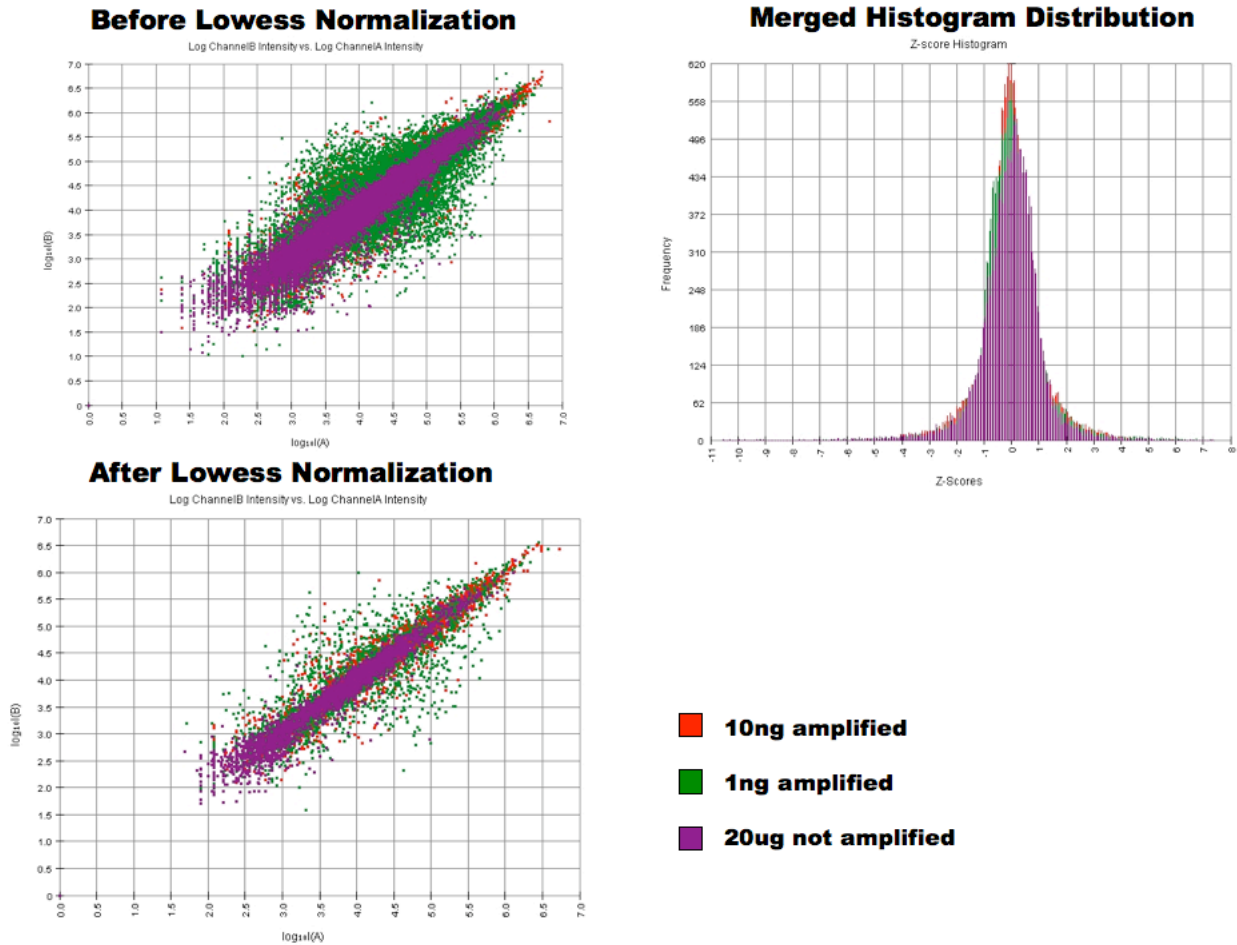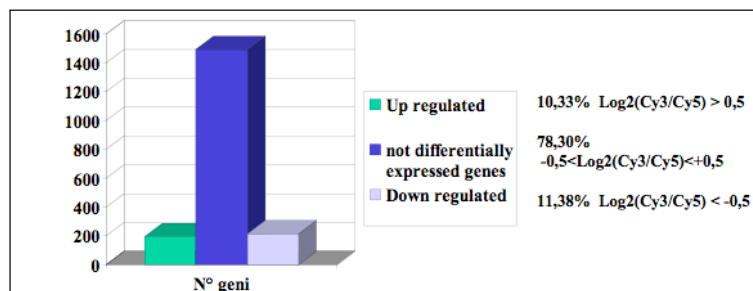
Figure 3-8. Scatter plot representing the distribution of all the signals detected by the microarray, calculated as the ratio of the logarithms of the signal intensities in the two separate channels (Cy3 and Cy5). The distribution is compared among the amplified samples (10 ng and 1 ng) and the unamplified sample (starting from 20 μg). It can be seen from the histogram distribution that there is a significant overlap. Only in the 1ng sample, before normalization is visible a more enlarged distribution after amplification, which is in part expected since amplification tends to amplify the relative differences. Although this could be acceptable untill the same trend in gene expression ratios is preserved. Reasonably a safe limit can be settled at 10 ng as minimal starting material for a more reliable preservation of the relative ratios of mRNAs.

The analysis of the data has been performed to assess the conservation in the relative ratios using the "true outliers" method, limiting to the differentially expressed genes and to the top candidates list (figure 3-9). As shown, the logarithms of the ratio of signal intensities in the two separate channels $Log_2(Cy3/Cy5)$ are significantly preserved for most of the outliers in the amplified samples as compared to the not amplified sample (20 μg). From all these data, we can conclude that the SMART7 amplification may effectively be used as a method of choice for gene expression profiling of small amounts of total RNA starting in the range of 1-10 ng.

Figure 3-9. Comparison among the lists of top ranked upregulated and downregulated outlier genes, amongst not amplified and amplified samples. As it can be seen from the values of the $Log_2(Cy3/Cy5)$ of the ratios of signal intensities in the two channels, the relative levels of the top differentially expressed genes are preserved at a good extent by the amplification process.

Developing of a new tecnique for Genome-Wide Tagging and Mapping of the Transcription Start Sites from limited amount of starting material: nanoCAGE and high-throughput sequencing

The classical CAGE protocol requires micrograms of RNA

Cap Analysis of Gene Expression (CAGE) has been establishing as the method for studying at a very large scale the distribution of the transcription start sites (TSS) and consequently the promoters usage in the genome (Shiraki T et al. 2003).

CAGE makes use of the cap-trapper technology (Carninci et al. 1996) to capture the 5` prime ends of the transcripts selecting the full-length 1st strand cDNA molecules. After the ligation of a

biotinilated 5`end linker to the single strand cDNA containing the recognition sites for two different restriction enzymes, XmaJI and MmeI, a $2^{nd}$ strand synthesis step is performed and the ds cDNA is digested by MmeI, which cuts 18-20 nucleotides downstream, generating short 5`end fragments, retained on streptavidin beads.

A second linker containing the restriction site for XbaI is ligated to the 3`end of these fragments and, after a PCR amplification step, the cleavage by XmaJI and XbaI generates 5` end tags, 20bp long, which can be ligated to each other to form concatamers and cloned into a vector for sequencing.

CAGE has been widely and successfully applied for the construction of hundreds of libraries from different mouse and human tissues, giving a great contribution to the understanding of the transcriptome structure.

In its original version, the CAGE protocol requires 25-50 μg mRNA as starting material, resulting in a big limitation in the scale to which it can be applied.

Although great efforts have been done to optimize each single step of the CAGE protocol, leading to sensible improvements of the method (Kodzius et al. 2006), the needs for large amounts of RNA as starting material makes this technique not suitable to investigate the transcriptome in the case of very small samples.

nanoCAGE – what is new: random priming, semi-suppressive PCR, starting from 10 ng

Here, we developed a completely new technique to apply CAGE to small scale samples. To this purpose, the initial "cap-trapping" step was substituted by taking advantage of a modification of SMART, a technique we previously used for small scale gene expression profiling.

The strand-switching properties of the reverse transcriptase were used to incorporate a small sequence tag in correnspondence of the cap structure of the transcripts during the 1st strand cDNA synthesis and then used to amplify selectively all the molecules that reached the 5'end. Priming

took place with a mixture of oligodT and random primer, ensuring that even the estimated 43% of the transcriptome which lacks a poly(A) tail (Chen et al 2005) may be studied.

A small single strand oligonucleotide, containing the recognition sites for Eco31I and the class IIs restriction enzyme EcoP15I, together with a terminal stretch of one dG and two riboguanosines in the 3'end, hav been designed to function in the strand-switching mechanism.

This oligonucleotide bears an additional sequence in the 5'end for PCR amplification which has no significant complementarity to the mouse and human genomes/transcriptomes as determined by running BLASTN 2.2.17 on the genomic and transcripts assemblies of those two species present in NCBI.

At first total RNA from whole body mouse embryo (C56BL/6 E17.5) was used as template. Serial dilutions provided scaling amounts of input RNA for a preliminary estimation of the lower limit of application of the technique.

In those first attempts, only directional priming by a modified oligodT, bearing a sequence employed for subsequent PCR amplification, was used.

Since random priming would result in a great improvement of the global representation of the transcriptome, we carried out several experiments to introduce this priming. However, our first attempts were not successful due to the massive formation of primer dimers which make artefactual products competing with the legitimate template in the PCR reaction.

Many different trials have been performed but without showing any significant improvement in reducing the formation of primer dimers (data not shown). Among them, we used short oligonucleotides complementary to the tail of the random primer; random primers containing a partial self-complementarity in the 5'end; a T7 endonuclease I, which recognizes and cleaves non-perfectly matched DNA sequences and heteroduplex DNAs or cruciform structures,

To overcome this problem, we decided to modify the design of the primers and to adopt a semi-suppressive PCR strategy, which basically takes advantage of a partial sequence complementarity among all the primers and makes possible to specifically inhibit the amplification of the byproducts

which are not fully complementary at the endpoints, avoiding the formation of artefacts in the PCR-mediated 2nd strand cDNA synthesis.

One of the major drawback of the template-switching technique is that the strand-switching oligonucleotide (TS-oligo) undergoes unspecific priming during 1st strand cDNA synthesis (Matz et al. 1999). Theoretically, in the process of the 1st strand synthesis, the strand-switching oligonucleotide may also function as RT primer, being extended in the 3'end, and the random primer could provide unspecific strand-switching activity at the 5'end as well: both of them will result in the formation of a single strand cDNA with a "panhandle-like structure" with fully complementarity at the endpoints.

Such kind of intramolecularly, paired structures would be more stable than the structures formed by the legitimate templates, which only display a partial complementarity at the endpoints.

The difference in the kinetics of fomation of more or less stable "panhandle-like structures" and of the corresponding relaxed structures can be used to selectively inhibit the PCR amplification of the byproducts taking place in the reverse transcription reaction.

A PCR-suppression effect (PS-effect) was described for the first time by Siebert et al. in 1995. When in complex samples containing short and long molecules with inverted terminal repeats (ITR) at the ends, after denaturation the single strand DNA tends to form a "panhandle-like structure" as a consequence of intramolecular pairing of the terminal ITRs. Importantly, shorter molecules are more easily suppressed than long fragments (Lukyanov et al. 1999; Shagin et al. 1999), because the intramolecular annealing is faster than the intermolecular annealing with the PCR primers. Since the intramolecular pairing is function of the length of the molecule, varying the Tm of the primers and the annealing temperature used for PCR would affect the proportion of molecules whose "panhandle-like structure" will have a kinetics compatible with the primer pairing in the PCR: on this basis one can modulate the size of the resulting amplification products in complex PCR mixtures, like full-length cDNA libraries.

In particular, four parameters are critical for the PS-effect: (1) the length of amplified DNA, (2) the concentration of primer, (3) the ratio of ITR lenghts to primer length (the difference in their melting temperature) and (4) the annealing temperature used in the PCR.

The suppressive PCR effect has been already used in previous studies (Siebert et al. 1995; Lukyanov et al. 1995; Diatchenko et al. 1996; Shagin et al. 1999; Piao et al. 2001). It proved to be effective in avoiding the formation of undesired short products during the amplification and to correct for the bias against long molecules which tend to be under-represented, due to the more efficient amplification of the shorter molecules in complex PCR.

The main difference between our strategy and the previously published studies is in the design of the primers and the terminal sequences used for tagging the cDNAs, which ensures that only the legitimate templates with a partially paired ends can be amplified but not the unspecifically primed products. Another possible approach to prevent the unspecific priming of the strand-swithcihng oligo (TS-oligo) would be to chemically modify this oligonucleotide in the 3'end to make it non extensible by the reverse transcriptase, as with the addition of a 3' phosphate group. However, in this case the problem of the unspecific cap-switch activity provided by the random primer may be more difficult to be solved.

The semi-suppressive PCR approach can circumvent both of these incovenients at the same time.

For the 2nd strand synthesis the number of cycles is particularly important: a small scale PCR reaction was performed to evaluate the optimal number of cycles, defined as the last cycle before the intensity of the products ceased to increase. Aliquotes (1/10 vol.) were taken every two cycles and analyzed on 1 % agarose gel.

The entire protocol has shown to be robust and highly reproducible, even starting from 1ng of total RNA, and even using partially degraded RNA obtained by laser capture microdissected cells, as shown by the pattern of the amplified cDNA in replicate experiments.

nanoCAGE-454 and nanoCAGE-Illumina-Solexa: new platforms for high-throughput gene expression profiling from small samples

The use of nanoCAGE has been coupled to high-throughput sequencing technologies such as 454 Life Sciences or Illumina-Solexa, to be effectively applied to a genome-wide scale from small samples and obtain a true full coverage of the transcriptome. It has been estimated that only sequencing millions of tags can ensure to cover all the different RNA species which could be present in a cell that are in the order of 400,000 or more.

More details about the specific results obtained by this new technique and its applications will be further illustrated in Section 4.

3.3 Application of the NanoCAGE technology to the mesencephalic dopaminergic neurons

LCM purification of A9 and A10 neurons and nanoCAGE amplification

A new technological improvement has been made to allow the Cap Analysis of Gene Expression (CAGE) technique to be applied to very small biological samples. The new CAGE-based technique was adapted for profiling gene expression starting from ng$_s$ of total RNA, hence named nanoCAGE. Briefly, Laser Capture Microdissection (LCM) was used by Christina Vlachouli to isolate 2000-3000 cells from each of two distinct populations of midbrain dopaminergic neurons, A9 and A10 respectively, from Th-GFP transgenic mouse, which express the Green Fluorescent Protein under the control of the rat tyrosine hydroxylase promoter. Total RNA extracted from these cells was retro-transcribed in the presence of a mix of oligodT and random primer (molar ratio 1:1) and a strand-switching oligonucleotide (riBOSS), which takes advantage of the cap-switching properties of the reverse transcriptase to specifically tag the 5`end of capped transcripts with a sequence

containing a class III restriction site for EcoP15I. The 1st strand cDNA was then amplified by means of a semi-suppressive PCR which minimize the risk to amplify primer dimers and other aspecific products (figure 3-10). The 2nd strand cDNA was purified on microcon filters to recover only the low molecular weight fraction containing the 5`tags. Then, the Illumina-Solexa adaptors, each pair containing a different dinucleotide barcode sequence (AC for A9, GA for A10), were ligated to the EcoP15I fragments. After another PCR amplification with low number of cycles (6-12 cycles), all the PCR products were separated on polyacrylamide gel and DNA (112 bp) was extracted, concentrated and sequenced with the Illumina-Solexa platform.
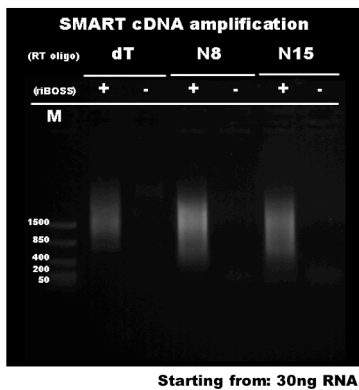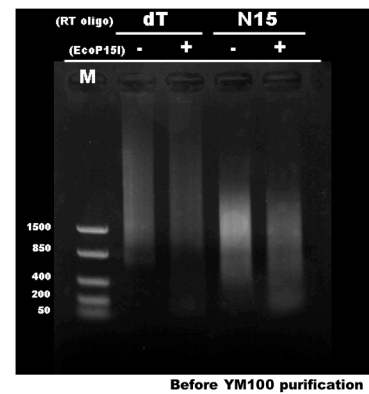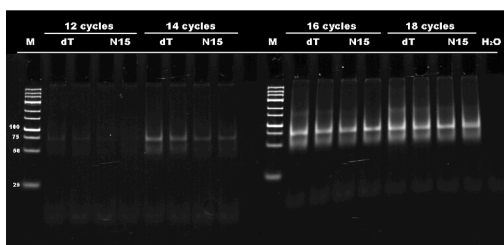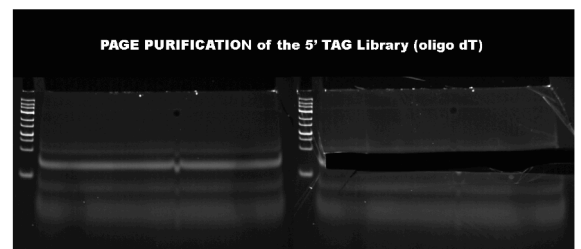
Figure 3-10. The various sequential steps necessary for 5`tags production in the nanoCAGE described in the text

By creating 32bp long 5`tags, we considerably improved the TSS mapping rate on the genome. The use of random priming in the 1st strand synthesis allowed to capture poly(A)$^{-}$ RNAs. The use of barcodes in the ligation step has the advantage to enable to mix different cDNA samples which in turn minimizes the bias introduced by the PCR in different reactions. Sequencing with Illumina-Solexa platform allows to have millions of sequence reads per experiment, making possible to reach an almost complete coverage of the transcriptome in a cell.
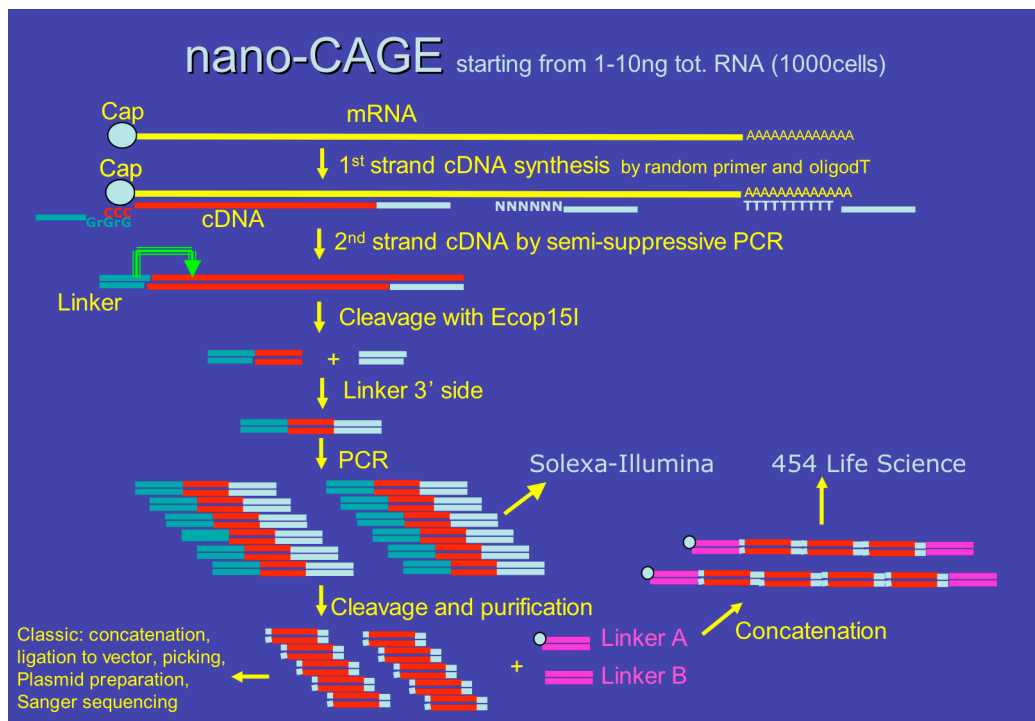
Figure 3-11. General scheme of the nanoCAGE procedure for library production

Bioinformatic analysis of sequencing data

The raw sequence data obtained from a full run (8 channels) of Illumina-Solexa sequencer was processed for "tag extraction" to discard unspecific sequences that did not contain the restriction sites used for tag production and the 2 nt barcodes inserted in the ligated oligonucleotides, to

distinguish among different samples in the mixed library. During tag extraction the first Gs were also removed as they came from the riBOSS strand-switching primer.

A statistical evaluation of the frequency of occurence of each of the 4 bases at each position of the tags has been done together with the construction of an error estimation model for better implementing the extraction process and lowering the fraction of false positives. Furthermore, a Pearson correlation coefficient was calculated for the sequence reads obtained from each of the 8 channels of the Solexa sequencer to assess if we could effectively mix all the sequences in the subsequent step of mapping and clustering without having any bias. Pearson correlation was quite high (0.991-0.998).

Here is an example of the Tag extraction process that has been applied to the raw sequences:

```
59045  GACCTTCAGTTCGACTACTGCTGGTACCCCTGC
49899  CTACCACTGAGCTAAATCCCCAACCCCCCCTGC
47199  CTTACCACTGAGCTAAATCCCCAACCCCCCCTG
44156  ACCCTTCAGTTCGACTACTGCTGGTACCCCTGC
21600  GAACCTTCAGTTCGACTACTGCTGGTACCCCTG
19704  CTCCTTCAGTTCGACTACTGCTGGTACCCCTGC
17717  ACTACCACTGAGCTAAATCCCCAACCCCCCCTG
14848  CTACCTTCAGTTCGACTACTGCTGGTACCCCTG
14468  ACACCACTGAGCTAAATCCCCAACCCCCCCTGC
14078  GATACCCCTGCTGGAGACCTTCAGTTCCCCTGC


59045  GA    CCTTCAGTTCGACTACTGCTGGTAC  CCCTGC
49899  CT    ACCACTGAGCTAAATCCCCAACCCC  CCCTGC
47199  CT    TACCACTGAGCTAAATCCCCAACCC  CCCCTG
44156  AC    CCTTCAGTTCGACTACTGCTGGTAC  CCCTGC
21600  GA    ACCTTCAGTTCGACTACTGCTGGTA  CCCCTG
19704  CT    CCTTCAGTTCGACTACTGCTGGTAC  CCCTGC
17717  AC    TACCACTGAGCTAAATCCCCAACCC  CCCCTG
14848  CT    ACCTTCAGTTCGACTACTGCTGGTA  CCCCTG
14468  AC    ACCACTGAGCTAAATCCCCAACCCC  CCCTGC
14078  GA    TACCCCTGCTGGAGACCTTCAGTTC  CCCTGC
```

Sequences start with GA, CT, or AC. These are the reverse-complement of the 2nt "barcodes" used for different DA-cell libraries (LC is AG, A10 is TC, and A9 is GT). All the reads finish with a given number of Cs, followed by TGC. This is due to the terminal addition of Cs to the first-strand cDNAs of capped transcripts, and CCCTGC is the reverse-complement of the 3' end of our strand-swiching oligo (riBOSS). Between these landmarks are the extracted tags.

After extraction was completed, the tag sequences were mapped to the mouse genome (assembly mm9) using KAlign software, and all the tags aligning with 100% identity to rDNA were discarded from subsequent analyses. The frequency of occurrence of each of the 4 nucleotides has been calculated for each position in the mapped tags to assess the quality of the data (Figure 3-12).

The catecholaminergic cell libraries were also compared to another library (OE olfactory epithelium), for which the same conditions of the nanoCAGE procedure were applied, and this was used as a useful outlier for statistical analyses.
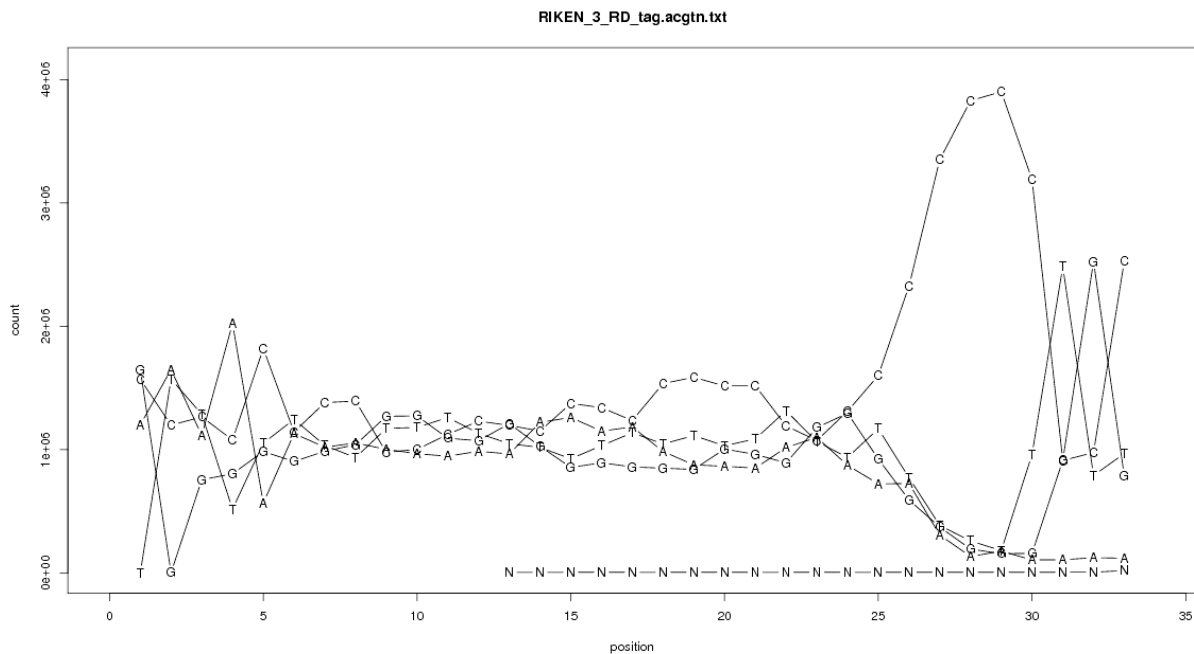
Figure 3-12. Distribution of the four bases at each nucleotide position along the nanoCAGE tags in A9-A10 libraries

As it can be seen in the following chart, all the libraries gave comparable ratios among total sequence reads, mappable tags and tags matching ribosomal DNA, even if they were processed independently and starting from different samples taken in independent LCM sessions.

nano CAGE library	CA library	OE library
total seq reads	36803476	21353318
total tags	28762748	18229100
ribosomal	1203752	232814
mapped tags	19946400	16568480

Figure 4-3. Numbers of the total sequence reads, tags and mappable tags aligned on the mm9 mouse genome

The tags mapping to rDNA were only a minority. Considering that uncapped ribosomal RNA represents more than 90% of the RNA present at steady state in a cell, this can be interpreted as a high success rate of the strand-switching mechanism to specifically targeting capped mRNAs.

185

nano CAGE library	tag counts	% over library
A10 single	4510297	31.9582506
A10 multi	8671523	61.4431168
A10 ribosomal	931271	6.5986324
A9 single	5313043	37.7262696
A9 multi	7587201	53.8743599
A9 ribosomal	1182895	8.3993703
OE single	13460466	41.2486824
OE multi	18686178	57.2625214
OE ribosomal	485831	1.488796
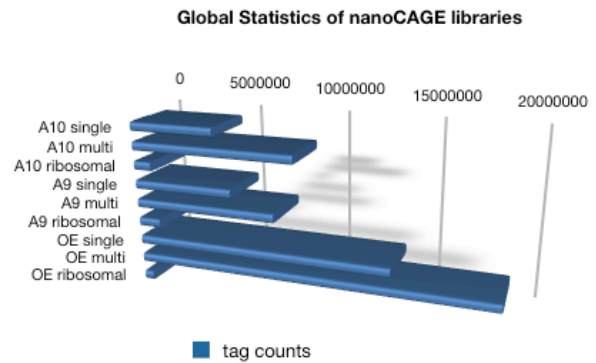
Figure 4-4. Tags counts for three libraries (A10, A9, OE) divided by mapping classes

As it can be seen in Figure 4-4, the percentage of the tags that mapped on more than one position in the genome (multimapping) was quite high. Dealing with multimapping CAGE tags rendered necessary to apply a special strategy for computing the associated tpm or tag count value. Briefly, if a tag match exactly more than one positon in the genome one can even decide to assign the value to that tag dividing it for the number of different mapped locations, or assign the value to the closest and bigger of the neighbour tags. The latter has been applied by Goef Faulkner who called it "guilty by association" method (Faulkner et al 2008).

Multimapping CAGE tags were rescued in a similar manner as the "guilty by association" method with the following modifications :

1.When a tag failed to be mapped onto the genome, NextAlign software then try to realign it allowing for one mutation or indel. This may give raise to tag multimapping (one mutated version aligning at a given position, while another mutated version align at another position). Those mutation are not equiprobable. The best 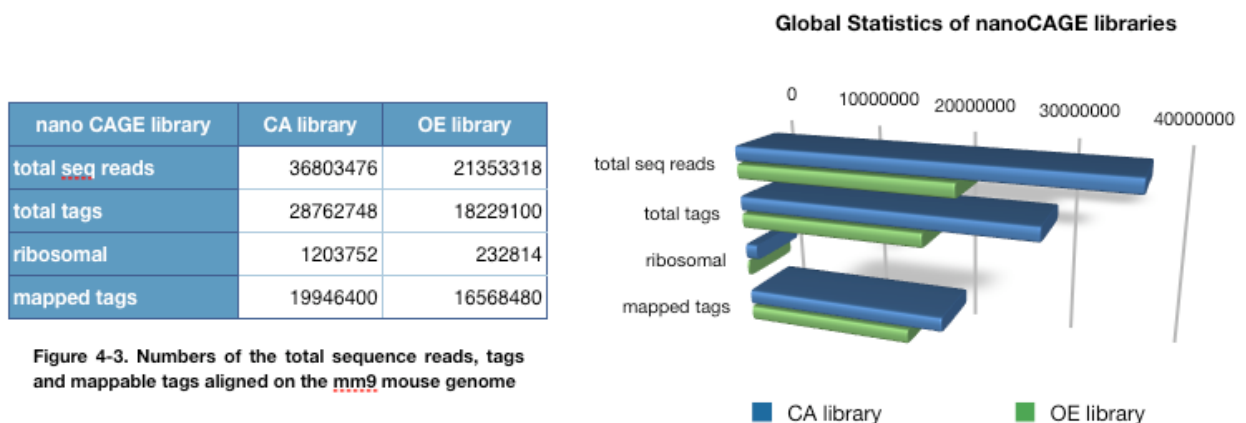example being the first G removal versus an indel within the tag. Goef's methods doesnt not take it into consideration, giving all mutated versions an equal probability to influence the guilty by association rescuing of multi mappers.

2.The weight given to each possible position of a multimapping tag is a function of the number of tags in a [-200bp..+ 200bp] window. Tags in close proximity to the position under consideration weight as much as tags distant of 100bp.

The tags distribution in the vicinity of a given tag (TSS) (considering both single mappers only and

186

single + multi mappers) was plotted as function of the distance, in order to model the influence of surrounding tags to weight potential positions of multimappers.

Because the frequency of tags at long distance is several order of magnitude smaller than tags in close proximity, a more permissive window of [-1kb..+1kb] was used to rescue mutlimappers. This yields a higher number of rescued tags, but also the inclusion of tags that do not have any in a [-200..+200 bp] window. An "ambiguity score" was computed which simply is the proportion of multi mapping tags (over all tags) for each TSS, which allow for further scrutiny when analyzing CAGE tag libraries and looking for novel TSSes.
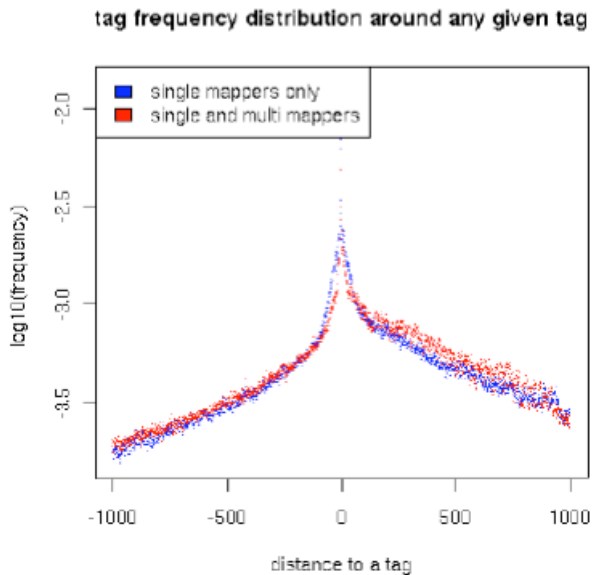
tag frequency distribution around any given tag

TSSes clustering strategies – (this part was contributed by Nicolas Bertin RIKEN,GSC, Japan)

Two strategies have been used to cluster TSSes : FANTOM3 style "proximity" tag clustering (TC) and Parametric clustering (PC).

In the FANTOM3 style "proximity" Tag Clustering" (TC), TSSes are clustered together if they are separated by 20 or less base pairs regardless of the RNA library they are derived, as in Carninci et al 2006. This clustering approach generated a total of 2,068,275 clusters (34% of the clusters with a tpm_density over 1) (Figure 3-13).

With the Martin Frith`s Parametric Clustering (PC) TSSes are clustered using the method described in Frith et al 2007. As noted by the author, a limitation of the proximity tag clustering approach is that TSS can only belong to one cluster while it is evident that one can observe the existence of clusters within clusters.

This approach enables a hierarchical definition of such clusters. In short, the algorithm employed

compute the local density (number of tags / length of the segment) of all genomic segments. The ratio of the minimum density value below which the cluster merges with another cluster and the maximum density value, above which the clusters splits into smaller clusters is used to score the cluster "stability" (if the ratio is small, then the cluster is not strongly present in the data and could easily vanish with some changes in the underlying tag counts, while if the ratio is big, the cluster is a prominent feature of the data).

This clustering approach generated a total of 4,736,538 clusters (526,461 clusters with a tpm_density over 1). It is important to note that those clusters are overlapping and that it generates clusters within clusters. A stability index bigger or equal to 2 has been chosen in Frith et al. to retain for relevant clusters. (all clusters regardless of their stability score are reported) (Figure 3-13)
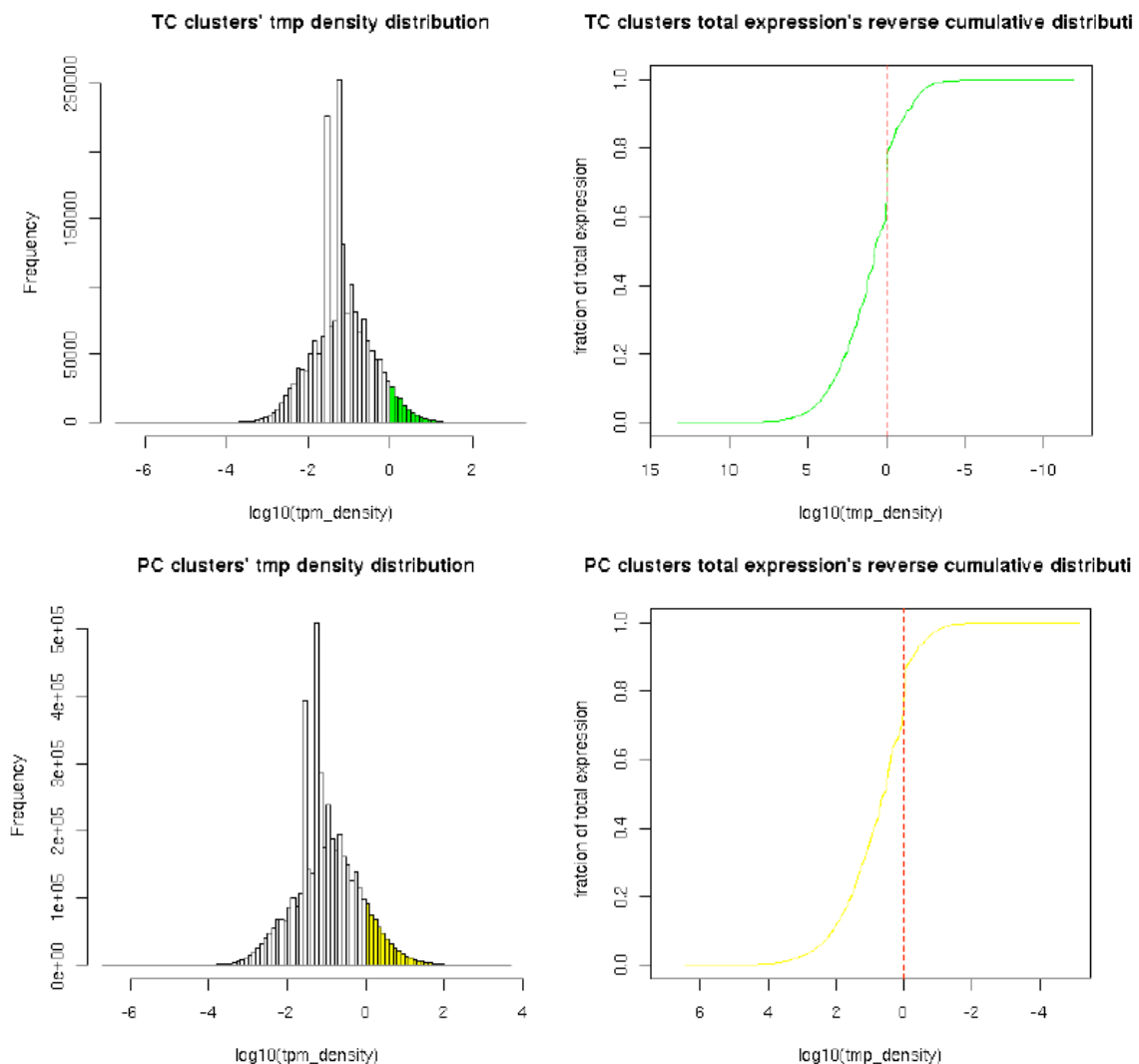
Figure 3-13. TC clusters tpm density distribution. The majority of the tags are associated to low expressed transcripts
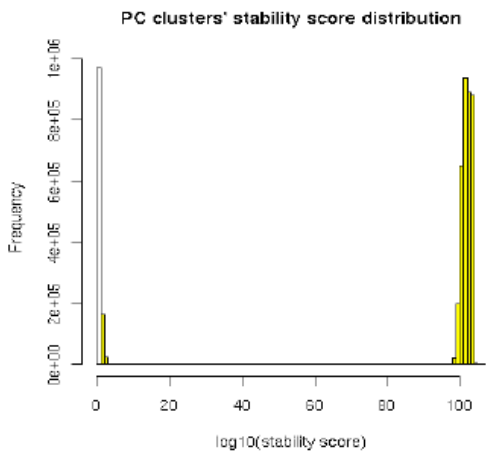
Figure 3-14. Stability scores of PC clusters are either low (<10) or extremely high (>100), 12% of PC clusters have a stability score < 2.

If we only retain clusters for which at least one tag has a non ambigous genomic location, ~34% of the TC clusters and ~42% of the PC clusters are retained as can be seen in figure 3-15.
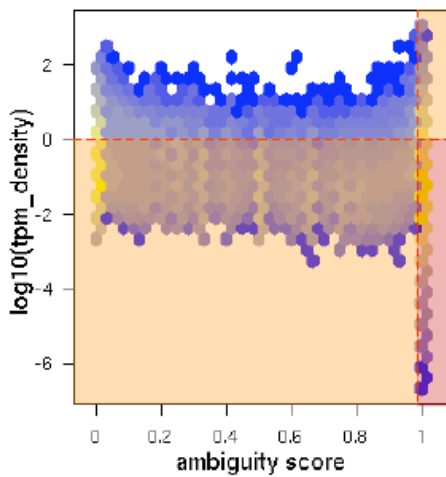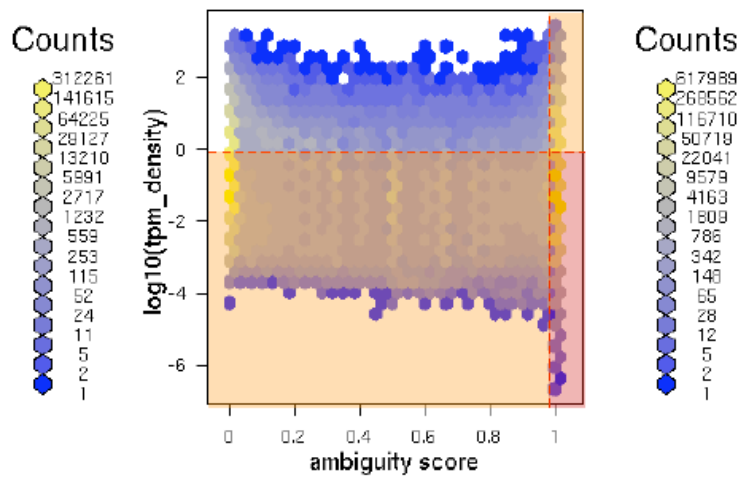
Figure 3-15. TC and PC clusters selection on the basis of the ambiguity score/density score ratio

Those 3 parameters (stability score, tpm density and ambiguity score thresholds) are not mutually exclusive.

Figure 3-16. A stability score/density score ratio showes that the majority of PC clusters falls into two main ranges, one enriched in highly stable clusters and another one enriched in very isolated tags. Both of these groups show a similar distribution of tpm density. The red dashed line represents the threshold over which the PC clusters have been retained.

189

Tag clusters-Transcripts associations

The genomic distribution along genes relative to the current annotations of the mouse genome has been calculated computing tag counts associated to each of the annotated RefSeq regions (exons, introns, proximal promoter and 3`UTR). As shown in figure 4-8 for A10 library and figure 4-9 for A9 library, there is a high percentage of TSSs in last exons, more than expected based on previous studies.

A10 single mappers	tag counts	% over library
last refseq exon S	1823553	35.1363495
middle refseq exon S	988861	19.0534444
first refseq exon S	700090	13.4893841
proximal promoter S	393857	7.5888648
middle refseq intron S	235230	4.5324284
first refseq intron S	105127	2.0255945
refseq distal 3UTR AS	72212	1.3913859
last refseq intron S	59022	1.1372401
last refseq exon AS	38442	0.7407032
middle refseq intron AS	36848	0.7099898
first refseq intron AS	28042	0.5403152
last refseq intron AS	13949	0.2687703
middle refseq exon AS	10245	0.1974013
first refseq exon AS	2699	0.0520045

Figure 4-8. Genomic distribution of A10 Tags classified according to their overlap with exons, introns, proximal promoters, 3`utr

A10 genomic TSS distribution

A9 single mappers	tag counts	% over library
last refseq exon S	1780117	29.2044077
middle refseq exon S	1166974	19.1452497
first refseq exon S	819236	13.4402975
proximal promoter S	569163	9.3376268
middle refseq intron S	392964	6.446925
first refseq intron S	186247	3.0555482
last refseq intron S	80059	1.3134393
middle refseq intron AS	62694	1.028551
refseq distal 3UTR AS	62219	1.0207582
last refseq exon AS	50606	0.8302365
first refseq intron AS	32043	0.525694
middle refseq exon AS	20506	0.3364192
last refseq intron AS	16846	0.2763736
first refseq exon AS	6398	0.1049648

Figure 4-9. Genomic distribution of A9 Tags classified according to their overlap with exons, introns, proximal promoters, 3`utr

A9 genomic TSS distribution

This 3`bias of the libraries could be in part explained by the usage of an equimolar mix of oligo dT

and random primer to prime the reverse transcription reaction. This effect could be due both to a more efficient priming of the poly(A)$^+$ RNAs by the oligodT, compared to a relatively less efficient priming of the random primer. On the other hand, the RNA quality may have also played a role, since total RNA obtained by LCM from limited amounts of samples has a lower quality than the RNA extracted and purified from abundant cultured cells. For partially degraded RNAs the use of random primer is needed.

However, alternative hypothesis may suggest that previous analysis may have underestimated the number of true promoters lying in 3'UTR regions of the genes. Carninci and co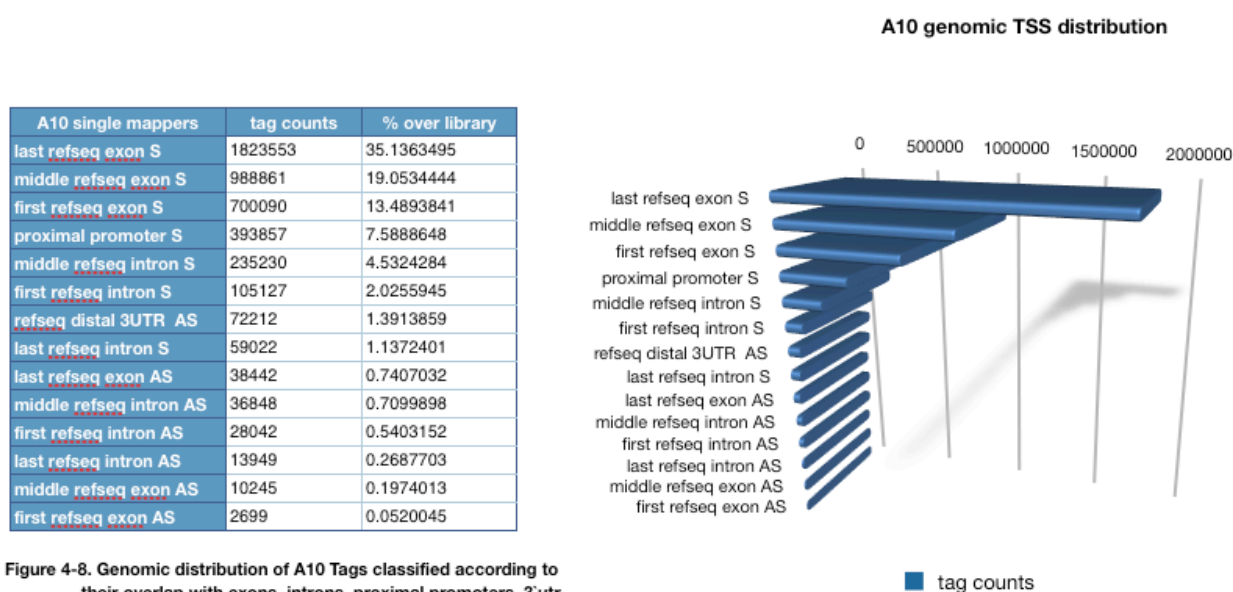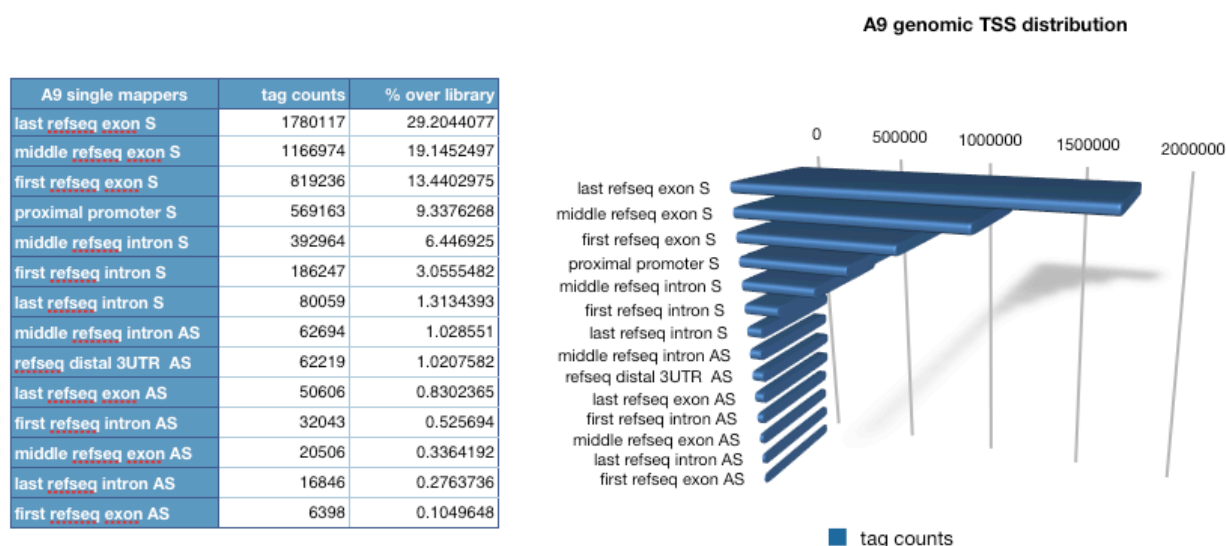lleagues have actually demonstrated that many promoters overlapping with 3`UTR regions guide the production of transcripts confirmed both by classical CAGE analysis and RACE experiments (Carninci et al 2006).

Then the informations about the tag clusters (single-mappers, multimapping, or both of them) were sorted based on the mouse genome annotations in three main categories: refseq genes, microRNAs and intergenic as it is summarized in the following Table 5.

rna_library	transcriptome_structure	co_direction	single_only_percents	multi_only_percents	single_multi_percents
A10	refseq	S	82.9633058	24.6103875	45.8213611
A10	refseq	AS	3.9005702	17.5516314	12.5895437
A10	mirna	S	0.2194633	0.0666817	0.122217
A10	mirna	AS	0.0020809	0.0004291	0.0010295
A10	intergenic	NA	12.914579	57.7708694	41.4658478
A9	refseq	S	81.9434942	24.5078814	49.4110755
A9	refseq	AS	4.1229973	14.8099013	10.1762248
A9	mirna	S	0.403913	0.1178742	0.2418962
A9	mirna	AS	0.0039538	0.0003767	0.0019277
A9	intergenic	NA	13.525641	60.5639655	40.1688749

For each cluster, the extent of the overlap with the transcript sets was extracted, as well as the overlap with region spanning [-500bp..-1bp] interval from the start site of each transcript annotation (later refer to as "proximal promoter"). Clusters that do not overlap with any transcript or their proximal promoter region have been annotated as "intergenic". Clearly, cluster transcript associations are not mutually exclusive; a cluster can be associated to more than one transcript, as in the case of overlapping genes. For each class data are as follows:

(1) REFSEQ genes set

20,649 refseq genes and their genomic coordinates were downloaded from UCSC mm9 release.

```
+----------------------------+----------+
| number_of_genomic_location | count    |
+----------------------------+----------+
|                          1 |    20536 |
|                          2 |       72 |
|                          3 |       19 |
|                          4 |        8 |
|                          5 |        3 |
|                          6 |        4 |
|                          7 |        2 |
|                          8 |        2 |
|                          9 |        1 |
|                         12 |        2 |
+----------------------------+----------+
```

They map to a total of 20,871 location on the genome. Break down of the multi-mapping refseq genes is illustrated in the Table 6.

Table6. RefSeq genes associated Tag counts distribution

(2) miRBase miRNA set

```
+----------------------------+----------+
| number_of_genomic_location | count    |
+----------------------------+----------+
|                          1 |      429 |
|                          3 |        3 |
|                          6 |        1 |
|                          8 |        1 |
|                         11 |        1 |
+----------------------------+----------+
```

Table 7. miRNA genes associated Tags counts distribution

436 miRNA annotated in miRBase and their genomic locations were retrieved from UCSC. Mature and precursor miRNAs from the miRNA Registry (miRBase) were aligned against the genome using blat. The extents of the precursor sequences were not generally known, and were predicted based on basepaired hairpin structure. miRBase is described in Griffiths-Jones et al 2006. The miRNA Registry is described in (Griffiths-Jones 2004 and Weber MJ 2005). Those 436 miRNA map collectively to 463 genomic locations. Break down of multi mapping miRNA is illustrated in the Table 7.

(3) Fantom3 ncRNAs set

```
+----------------------+-------------+
| subclass             | ncRNA_count |
+----------------------+-------------+
| F3                   |        2873 |
| F3_vote              |       35040 |
| F3_conservative      |         766 |
| F3_conservative_vote |        2881 |
+----------------------+-------------+
```

Table 8. Fantom3 ncRNA associated Tag counts distribution

Noncoding RNA sequences were collected from the FANTOM3 ftp site (ftp://fantom.gsc.riken.jp/FANTOM3/noncoding) and realigned onto mm9 using criteria similar to that used by UCSC blat based aligment pipeline. Sequence alignement against the mouse genome were performed by using blat. Those with an alignment of less than 15% were discarded. Furthermore, when a single sequence aligned in multiple places, the alignment having the highest

192

base identity was chosen. Only alignments having a base identity level within 0.1% of the best and at least 96% base identity with the genomic sequence were kept.

The FANTOM 3 ncRNA dataset is composed of the following 3 overlapping subsets :

- **fantom3_noncoding (F3)** including 34,030 potentially noncoding transcripts in FANTOM 3 clones, as determined by manual annotation.
- **f3_ncRNA_conservative** including 3,652 FANTOM3 clones that are annotated as noncoding, and in addition have experimental support for completeness of their 5' and 3' ends, and are not upstream of A-rich tracts in the genome. This set was made by intersecting f3_ncRNA_liberal with f3_mm5_fulllength in the artifact/truncation directory.
- **f3_noncoding_vote** comprising 38,129 FANTOM3 clones predicted to be noncoding by a majority vote of CRITICA, mTRANS040701, and rsCDS gene structure prediction tools.
-

Our most stringent set of 2,881 ncRNAs can be obtained by intersecting f3_ncRNA_conservative with f3_noncoding_vote (F3 conservative vote in Table 9).

```
+----------------------+---------------------------+--------+
| subclass             | number_of_genomic_location | count |
+----------------------+---------------------------+--------+
| F3                   |                         1 |  2752 |
| F3_vote              |                         1 | 34166 |
| F3_conservative      |                         1 |   757 |
| F3_conservative_vote |                         1 |  2848 |
| F3                   |                         2 |    75 |
| F3_vote              |                         2 |   583 |
| F3_conservative      |                         2 |     6 |
| F3_conservative_vote |                         2 |    26 |
| F3                   |                         3 |    17 |
| F3_vote              |                         3 |   100 |
| F3_conservative      |                         3 |     3 |
| F3_conservative_vote |                         3 |     2 |
| F3                   |                         4 |     2 |
| F3_vote              |                         4 |    41 |
| F3_conservative_vote |                         4 |     3 |
| F3                   |                         5 |     3 |
| F3_vote              |                         5 |    21 |
| F3_conservative_vote |                         5 |     1 |
| F3                   |                         6 |     2 |
| F3_vote              |                         6 |    15 |
| F3_conservative_vote |                         6 |     1 |
| F3                   |                        >6 |   114 |
| F3_vote              |                        >6 |    22 |
| F3_conservative_vote |                        >6 |     0 |
+----------------------+---------------------------+--------+
```

Table 9. Tag counts distribution across the ncRNA subclasses.
Those 41,560 mcRNA map collectively to 47,967 genomic locations.
Break down of multi mapping miRNA is as follow

In order to pinpoint interesting validation candidates, mm9 phastcons annotations from UCSC were retrieved. Predictions of conserved elements produced by the phastCons program are based on a whole-genome alignment of vertebrates, the placental mammal subset of species in the alignment, and the euarchontoglire subset of species in the alignment. They are based on a phylogenetic hidden Markov model (phyloHMM), a type of probabilistic model that describes both the process of DNA

substitution at each site in a genome and the way this process changes from one site to the next.

Only the "Vertebrates set" phastcons scored segments. They were retrieved as corresponding to the most conservative set (30 species all the way down to fugu).

For each cluster, the mean and maximum phastcons score (allowing to quickly determine when the cluster spans on several phastcons segments with different score) were computed. A score 3 of 1 means that the cluster seats on a genomic segment that is fully conserved down to fugu, a score of 0 means that this genomic region isn't conserved.
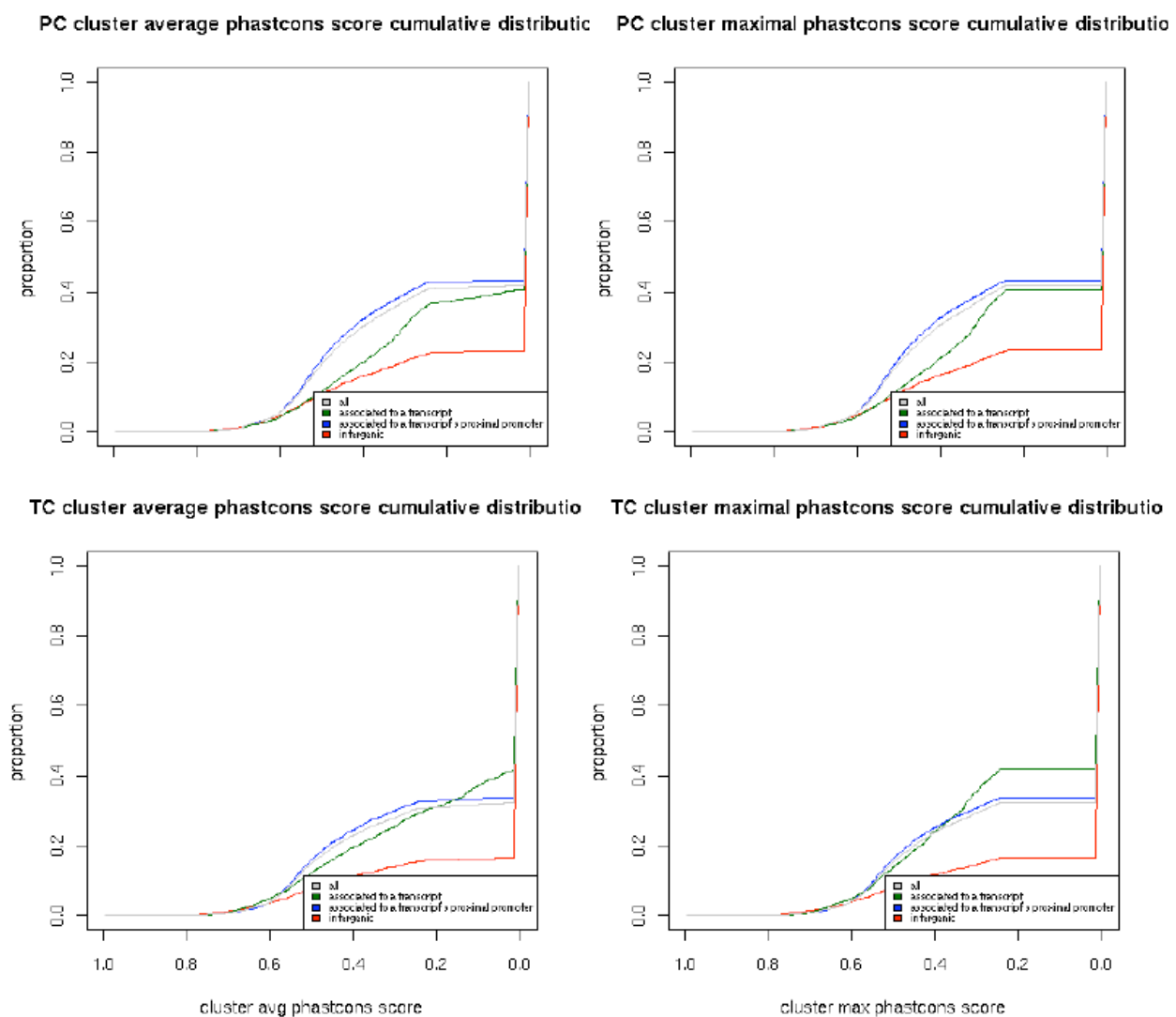

Figure 3-17. Distribution of the degree of conservation of PC and TC tag clusters

There aren't any clusters (PC or TC) completely conserved down to fugu (maximum phastcons score of 0.817). However, PC clusters associated to proximal promoter regions are more conserved than clusters associated to transcripts or intergenic regions. Whereas TC clusters associated to

transcript are more conserved than those associated to transcripts or intergenic regions. It might be important to keep in mind the hierarchical nature of PC clusters (existence of clusters within clusters) in interpreting these results. Imposing at least some level of conservation would get ride of ~60% of the clusters

Taking the maximum scoring segment or the average didn't seem to yield a very different outcome for clusters associated with proximal promoter or intergenic regions. Whole transcript exon painting might explain the difference between TC cluster avg phastcons score distribution and TC cluster max phastcons score distribution.

Furthermore, an analysis of the clusters length distribution reveals that about 14% of PC clusters and about 31% of TC clusters, respectively, are longer than 1bp. Clusters associated to transcripts are longer than those associated to promoters or intergenic regions.
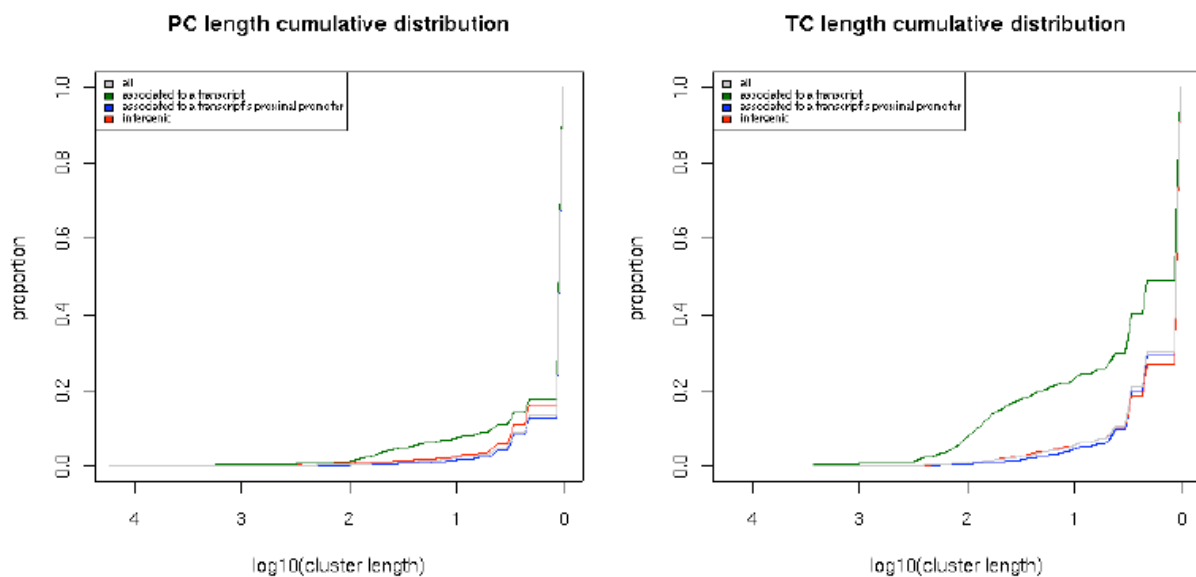

Figure 3-18. Clusters length distribution

This may reflect the so called "exon painting", probably due to the presence of functional promoters in internal exons of known genes. This phenomenon has been already described (Carninci et al 2006). This result could be particularly striking since the exon painting was originally observed in classical CAGE libraries which used a completely independent method for 5`tags production.

Generation of a list of potential target genes for further experimental validation

For the validation of potential target genes, all the informations detailed above has been gathered in 2 tab text files (one for each PC and TC clustering methods), together with the expression level for each of the A9, A10 libraries, and conservation, stability (PC only) and ambiguity scores.

Entries in both tab delimited files are ordered in decreasing values of A9/A10 expression ratio to favor the quick retrieval of A9 but non A10 expressed clusters.

One line for each unique cluster / position / associated transcript
And contains the following columns :
[1] "cluster_id" : internal cluster_id
[2] "chr" : coordinate of the cluster (in the form chr1, chr2, ...chrX, chrY, chrM)
[3] "strand" : coordinate of the cluster (one of "-, '+')
[4] "start" : coordinate of the cluster (start position of cluster with start < end systematically)
[5] "end" : coordinate of the cluster (end position of cluster with end > start systematically)
[6] "length" : the length of the cluster
[7] "avg_phastcons_score" : the mean phastCons score of the segments overlapping the cluster (in the range [0..1])
[8] "max_phastcons_score" : the phastCons score of the highest scoring segment overlapping the cluster (in the range [0..1])
[9] "ambiguity_score" : the ratio of multimapping position within the cluster (in the range [0..1])
[10] "stability_score" : for PC cluster only (blank for TC cluster), higher the value, more "stable" the cluster
[11] "has_associated_transcript" : one of
   • 'transcript'; the cluster overlap at least in part with a refseq gene, miRNA or ncRNA
   • 'proximal_promoter'; the cluster overlap at least in part with the proximal promoter region ([500.. 1bp]) of a refseq gene, miRNA or ncRNA

   • 'proximal_promoter,transcript'; the cluster overlap at least in part with a refseq gene, miRNA or ncRNA and in part with a refseq gene, miRNA or ncRNA proximal promoter region (remark : not necessarily the same transcript in principle ... in practice, this is very likely)
   • 'intergenic'; the cluster does not overlap with any transcript or any proximal promoter
[12] "transcript_name" : the name of the refseq gene, miRNA or ncRNA overlapping with the cluster (blank if the cluster map to an intergenic region)
[13] "transcript_annotation" : one of 'mRNA', 'miRNA' or 'ncRNA' (blank if the clustre map to an intergenic region)
[14] "annotation_origin" : one of 'refseq', 'mirbase', 'F3', 'F3_vote', 'F3_conservative' or 'F3_conservative_vote' (blank if the cluster map to an intergenic region)
[15] "co_direction" : one of 'S' : the cluster map on the same strand as a transcript 'AS' : the cluster map on the opposite strand to a transcript; blank if the cluster map to an intergenic region
[16] "association_span" : the relative position of the cluster start to the transcript or proximal promoter of the transcript (ranges from [1.. 0] when mapping to a proximal promoter and [0..1] when mapping to a transcript)
[17] "association_span" : the relative position of the cluster end to the transcript or proximal promoter of the transcript (ranges from [1.. 0] when mapping to a proximal promoter and [0..1] when mapping to a transcript)
[18] "A9" : cluster tpm (sum over all TSSes) for the A9 library
[19] "A10" : cluster tpm (sum over all TSSes) for the A10 library
[20] "OE" : cluster tpm (sum over all TSSes) for the OE library

```
"cluster_id"    "chr"   "strand""start"         end"      "length"          "avg_phastcons_score"   "max_phastcons_score"   "ambiguity_score"          "st
"10"    4658970 "chrX"  "-"     71032217        71032248        32      0.0946875       0.303   0       2.01900402273552        "proximal_promoter,
"11"    4658981 "chrX"  "-"     71032190        71032309        120     0.0781785714285714      0.337   0       5.91106624615927        "proximal_p
"16"    4658969 "chrX"  "-"     71032221        71032248        28      0.108214285714286       0.303   0       8.74066835693541        "proximal_p
"20"    171173  "chr1"  "-"     29542770        29542781        12      0.425444444444445       0.547   0.251207729     2.78682022336253        "in
"23"    171169  "chr1"  "-"     29542779        29542780        2       0.2735  0.547   0.1125  3.46819112589459        "intergenic"    ""      ""
"24"    171168  "chr1"  "-"     29542780        29542780        1       0       0       0.04054054      8.1580688881522e+98     "intergenic"    ""
"30"    1798102 "chr17" "+"     25716109        25716168        60      0.0187142857142857      0.262   0.004830917     11.8311011268647        "pr
"32"    1798098 "chr17" "+"     25716110        25716164        55      0       0       0.004950495     5.04462323102963        "proximal_promoter,
"33"    1798097 "chr17" "+"     25716111        25716164        54      0       0       0.005025125     3.41921560650376        "proximal_promoter,
"34"    4658953 "chrX"  "-"     71032222        71032222        1       0       0       0       1.49667278688251e+99    "transcript"    "NM_008323"
"41"    412823  "chr10" "+"     127523145       127523157       13      0.597   0.597   0.020134228     2.15455572391635        "transcript"    "NM
```

Figure 3-19. Example of how appears the header of those Table files

Starting from those tables, a first screening was done looking at the ratio of expression in A9 versus all (sum) other libraries. In figure 4-13 the x-axis is the log10(tpm) for A9, whereas the y-axis represents the ratio tpm(A9) / total tpm(A9, A10, OE) plotted on an exponential scale to emphasize larger, more interesting, ratio values.
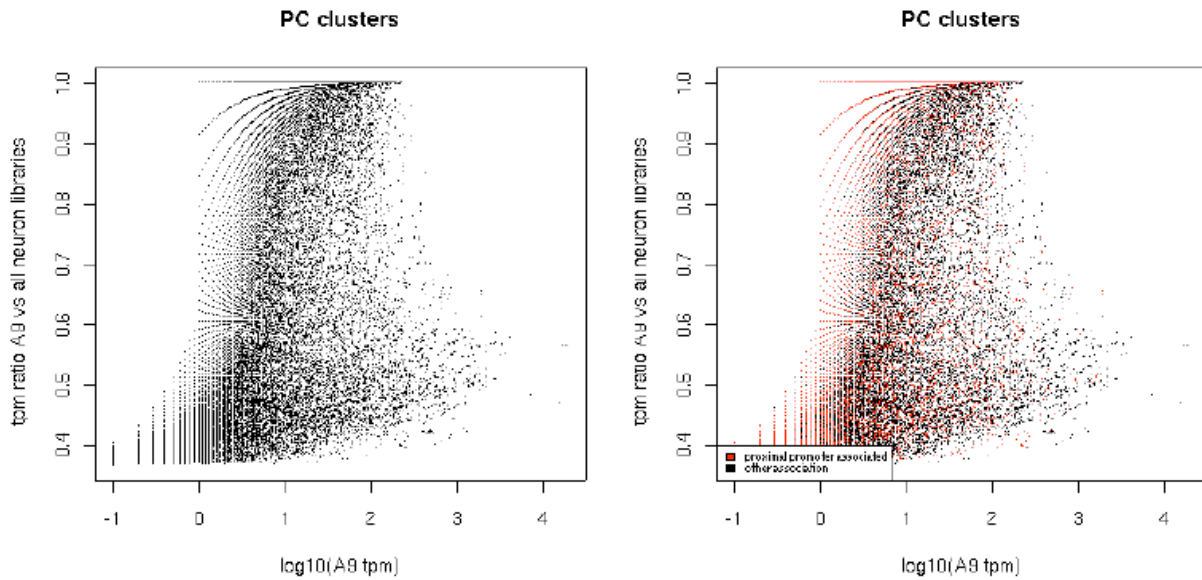

Figure 3-20. Distribution graph of A9 tpm density versus all other libraries tpm, in log scale. In red are represented clusters associated to proximal promoters, in black are the clusters associated to all toher annotaions

Restricting the analysis to only proximal promoters-associated PC clusters, similar results were obtained. In the following figure, clusters are also colored according to max-phastcons-score.
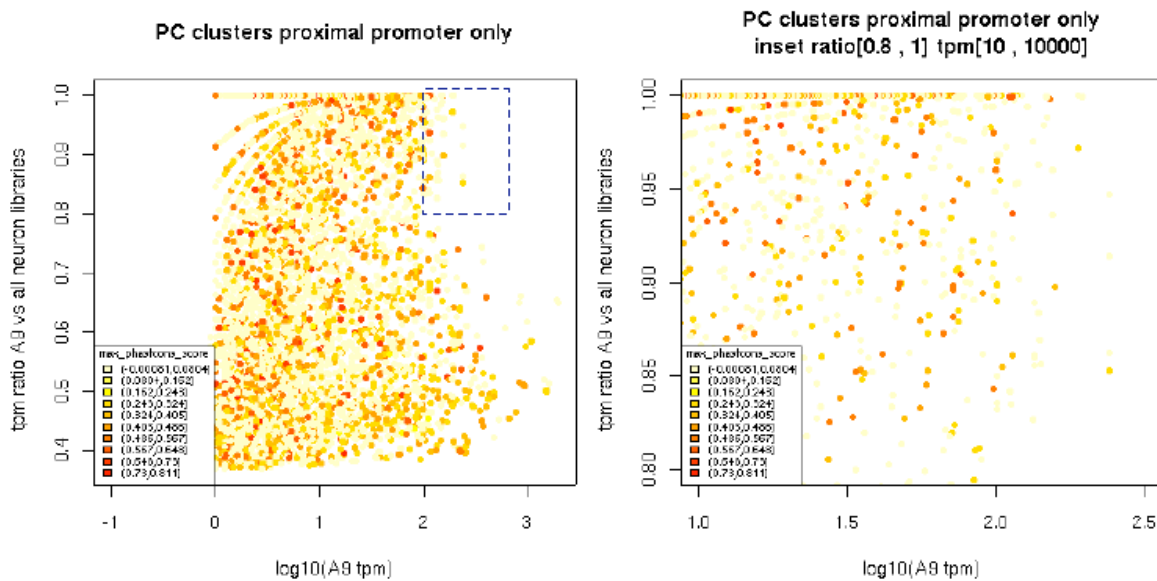

Figure 3-21. Distribution graph of proximal promoter associated PC clusters sorted by A9/all density and conservation

197

The same kind of analysis have been done selecting the PC clusters on the basis of annotation type
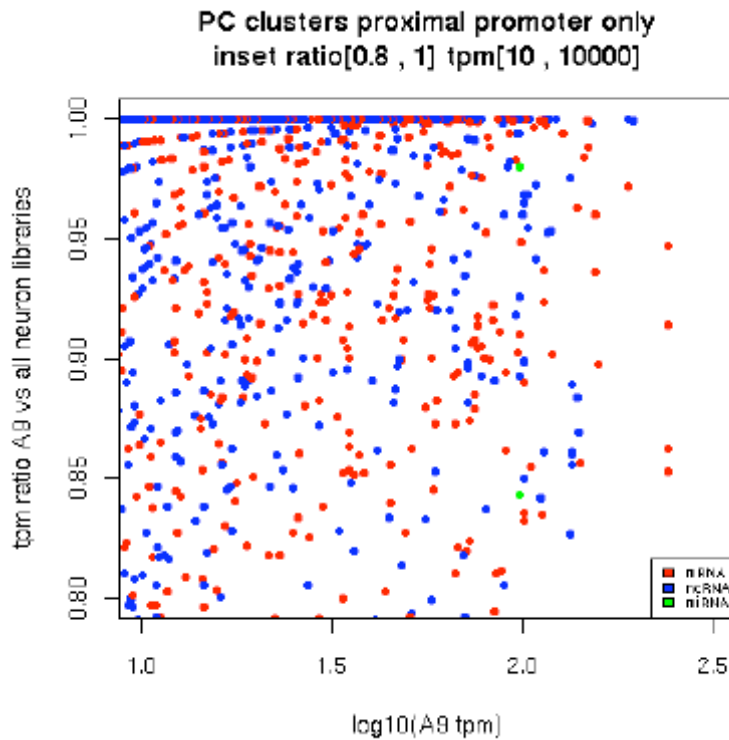

Figure 3-22. Distribution graph of proximal promoter associated PC clusters sorted by A9/all tpm density and color – coded according to different annotation types. Protein coding mRNAs (red) ncRNAs (blue) miRNAs (green).

GO terms and categories distribution

In order to look for GO terms categories distribution among A9 and A10 nanoCAGE clusters associated RefSeq genes, a first attempt was performed with FatiGo+ bioinformatics open-source tool (Al-Shahrour et al 2007). This did not detect any significant enrichment for any specific category.

One limitation in GO terms analysis is that only protein-coding genes can be considered. Moreover also many of the coding genes still lack any GO associated classification.

But it has to be noted that with nanoCAGE tags one have to deal with the additional information about the precise location of the tags along the transcripts. So if only proximal promoter associated tags are considered, then the number of different genes in the analysis can become a limiting factor.

To have a raw idea of which are the most expressed genes among A9 and A10 cells, a selection has been done using R programming, sorting the clusters on the basis of their tpm values in crescent order in each library. Herein after are the top 10 most expressed clusters detected in A10 and A9 respectively (Table 10).

```
        cluster      A9        A10                    cluster      A9        A10
Ndufa3     3613 6640.1392 2702.406    Atp5l    206405 2874.755 1751.162
Uchl1    312600  778.8844 2772.046    Ndufs7   379281 2994.925 1904.370
Cst3     649917 1592.3563 2953.486    Psmb6    328580 3314.416 1623.928
Rpl38    542374 2343.7481 3268.184    Cox7a2   378612 3614.396 3843.937
Ndufa13  339130 5851.2776 3286.817             432131 3833.672 1994.902
         359862 1896.7709 3753.216    Arl21     19591 3916.372 1506.670
Cox7a2   378612 3614.3961 3843.937    Ndufa4    43203 4752.237 4852.210
Syp       24826 8572.6062 4320.123    Ndufa13  339130 5851.278 3286.817
Ndufa4    43203 4752.2369 4852.210    Ndufa3     3613 6640.139 2702.406
Slc25a4  218970 2700.2656 5159.567    Syp       24826 8572.606 4320.123
```

Table 10.   Top 10 clusters in A10 library   Top 10 clusters in A9 library (values are in tpm)

From this top 10 list emerges that metabolism related genes seem to be very important for the function of these neurons.

A more time-consuming analysis was done by visual inspection of the values and positions of the tag clusters for entire sets of genes, classified based on common Pathways (retrieved from KEGG, Common Pathways, Ingenuity, BioCarta, GeneNet), using nanoCAGE custom tracks in wig format uploaded on the UCSC Genome Browser. Moreover, for each gene present in the pathway, additional informations in the literature were retrieved and screened either using the automated PubMed links present in UCSCS Genome Browser or consulting other sources of informations, such as GeneCards, Gene Profiles, or other specific databases for specific categories of genes.

First of all, some basic controls were performed looking at the expression of the genes involved in the catecholamine synthesis (figure 3-23) pathway, dopamine degradation pathway, and dopamine uptake and signalling pathways (figure 3-24). Furthermore, a series of other controls were performed on the quality of the data, looking for the expression of genes which are considered markers of other neuronal cell types including, among others, Dbh for NA-ergic neurons, GFAP for glial cells and Ermap for erythoid cells. Almost all of these control marker genes have no
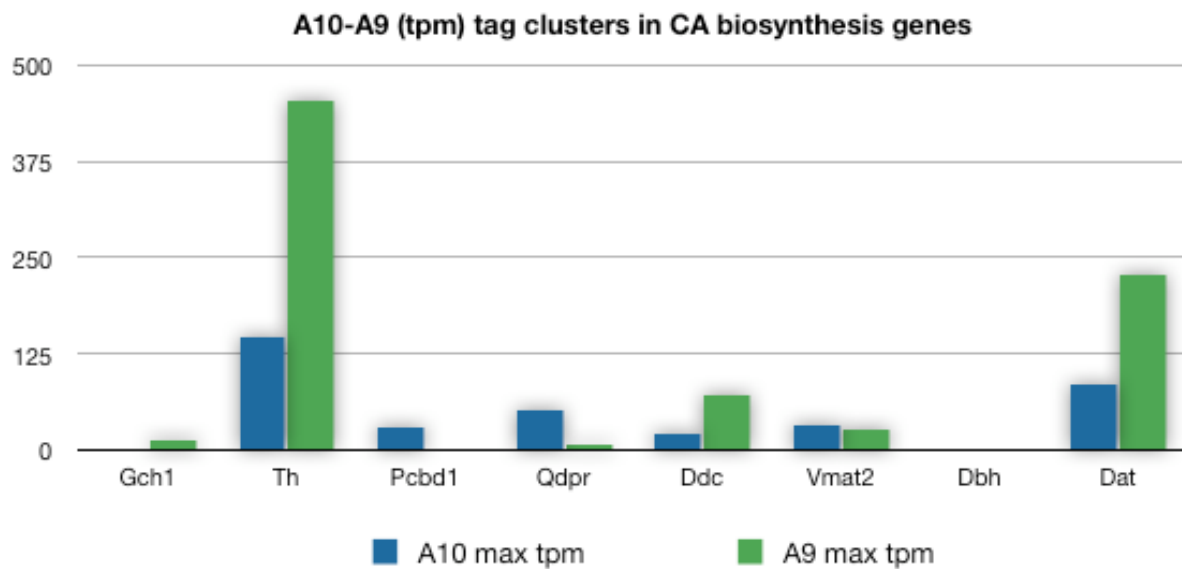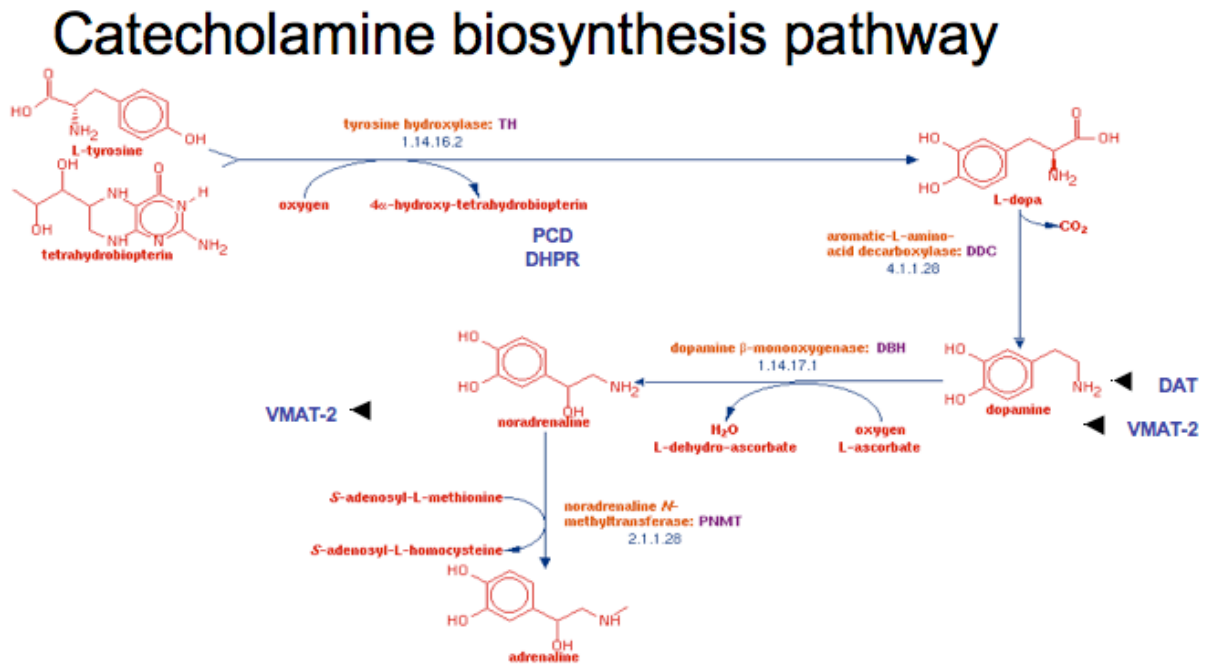
associated tag clusters in A9 and A10 DA cell libraries.

Figure 3-23. Catecholamine biosynthesis pathway genes are all represented in A9 and A10 libraries
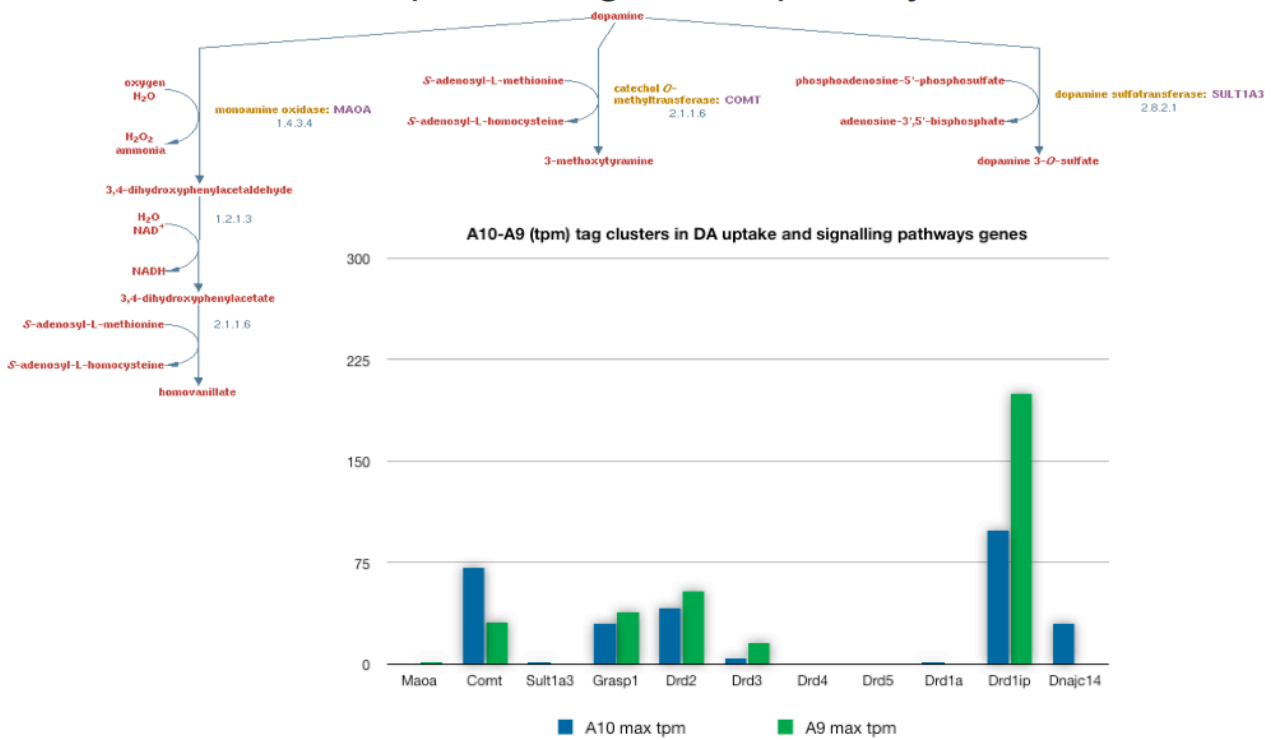
Figure 3-24. Dopamine degradation pathway genes associated tag clusters are epresented for A9 and A10 libraries

Furthermore, we examined the information about main pathways already known to be implicated in Parkinson`s disease, such as mitochondrial function, ubiquitin-proteasome system related genes, and the Parkinson`s disease pathway from KEGG. For many genes in those pathways a significant difference was present in the expression (max tpm) amongst A9 and A10 libraries.

Here there is a table with all the tpm values of clusters associated to the genes in the Parkinson`s Pathway as rom KEGG (Kyoto Enciclopedia of Genes and Genomes) table11

kegg_	Gene _name	refseq_name	proximal_A9_tpm	proximal_A10_tpm	distal_A9_tpm	distal_A10_tpm
Parki	DJ-1	NM_020569	14.37 [2.75-79.11]	36.26 [3.91-360.96]	55.71	65.13
Parki	Ube2j1	NM_019586	0	0	34.3	73.48
Parki	Ube1y1	NM_011667	0	0	0.04	0.01
Parki		XM_001474599	NULL	NULL	NULL	NULL
Parki	Pink1	NM_026880	0	0	54.77	57.17
Parki	Park2	NM_016694	0	0	31.72	36.32
Parki	Sncaip	NM_026408	0	0	12.6	30.51
Parki	Gpr37	NM_010338	0	0	0.54	10.84
Parki	Th	NM_009377	30.59 [3.67-273.12]	67.78 [4.96-1006.95]	2587.61	1241.91
Parki	Sept5	NM_213614	0	0	34.01	44.89
Parki	Ube2j2	NM_001039157	0	0	14.21	4.1
Parki	Stx1a	NM_016801	0	64.89 [4.88-937.50]	1.42	3.78
Parki	Ubb	NM_011664	0.04 [0.29-0.01]	16.71 [2.92-101.32]	666.94	795.89
Parki	Uchl1	NM_011670	267.36 [8.36-9559.57]	1164.63 [14.63-106794.35]	1563.95	1308.62
Parki	Ube1l	NM_023738	0		0	0
Parki	Ube2l6	NM_019949	0	66.98 [4.94-987.53]	0.63	1.16
Parki	Ube1x	NM_009457	0	0	198.38	162.24
Parki	Snca	NM_001042451	0	0	382.25	232.83
Parki	Ube2g2	NM_019803	0	0	58.25	3.09
Parki	Ube2g1	NM_025985	0	0	30.59	30.13
Parki	Ube2l3	NM_009456	0	4.51 [1.77-11.83]	62.36	82.71

Table 11. Parkinson`s disease KEGG Pathway associated genes- the Expression associated to a given transcript has been splitted into proximal (promoter and 1st half of the transcript genomic span) and distal (last half of the transcript

201

genomic span). Numbers in bracket are the [-2stddev +2stddev] I mentioned earlier.

Differentially expressed protein-coding transcripts in SN and VTA neurons

In general, many interesting candidate genes for further validation were found among those implicated in the DA metabolism, uptake and recycling. Some of these differ only in the relative level of expression. Others may also show alternative TSSs, hence possibly controlled by cell-type alternative specific promoters. Some of these protein encoding genes have been selected as targets for further experimental validation and will be object for discussion later. Other candidate genes for validation have been also found implicated in the mitochondrial homeostasis, regulation of mitochondrial fission and fusion dynamic process, and among those implicated in the function of the ubiquitin-proteasome system. Herein are the statistical analysis of the nanoCAGE clusters associated to genes implicated in crucial aspects of mitochondrial functions (Figure 3-25).
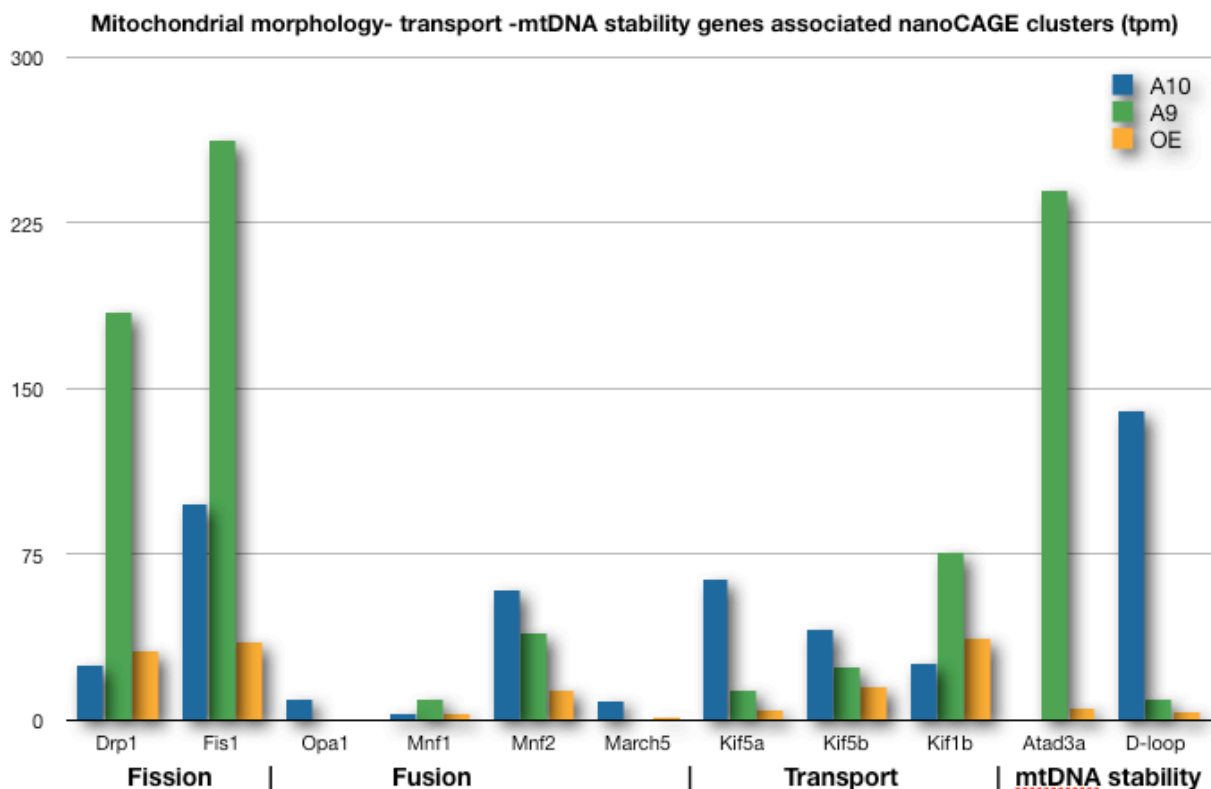

Figure 3-25. distribution of the nanoCAGE tag clusters (expressed as tpm) among three libraries, two A9, A10 and the olfactory epithelium (OE) library used as reference for comparison.

The genes represented in figure 3-25 are sorted in four functional categories: Drp1 and Fis1 are implicated in the control of mitochondrial fission, which is suspected to be involved in critical aspect of the neurodegeneration which affect the DA SN neurons in PD. Recently it has been shown that 6-hydroxydopamine could induce mitochondrial fission (fragmentation) in SH-SY cells in a Drp1-dependent manner (Gomez-Lazaro et al 2008). Moreover Pink1, implicated in familial forms of PC, has been shown regulate the mitochondrial dynamics in vivo interacting with the fission-fusion machinery in Drosophila and in mammalian cells (Yang et al 2008).

An RT-PCR analysis was done to confirm the expression of the specific genes of the mitochondrial fission-fusion apparatus (figure 3-25), using total RNA extracted from ventral midbrain of C57Bl6 adult mice and all the candidate genes (Drp1, Fis1, Opa1, March5) were confirmed to be expressed. We are currently performing real time PCRs on LCM-purified A9 and A10 neurons to confirm the differential expression of these genes. These data may suggest a different mitochondrial fusion/fission homeostasis between A9 and A10. This may have some consequences on the differential susceptibility of these neurons in PD.

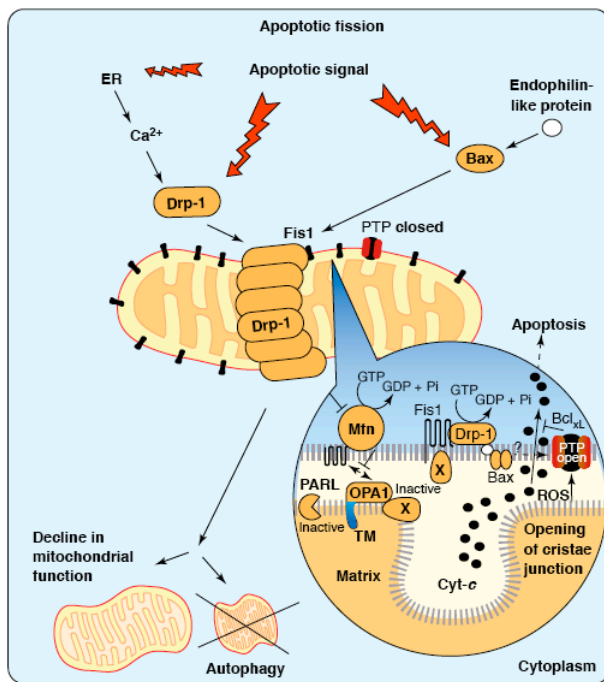**Figure 3-26.** Schematic draws of mitochondrial fusion and fission pathways – from (Bossy-Wetzel et al 2003)
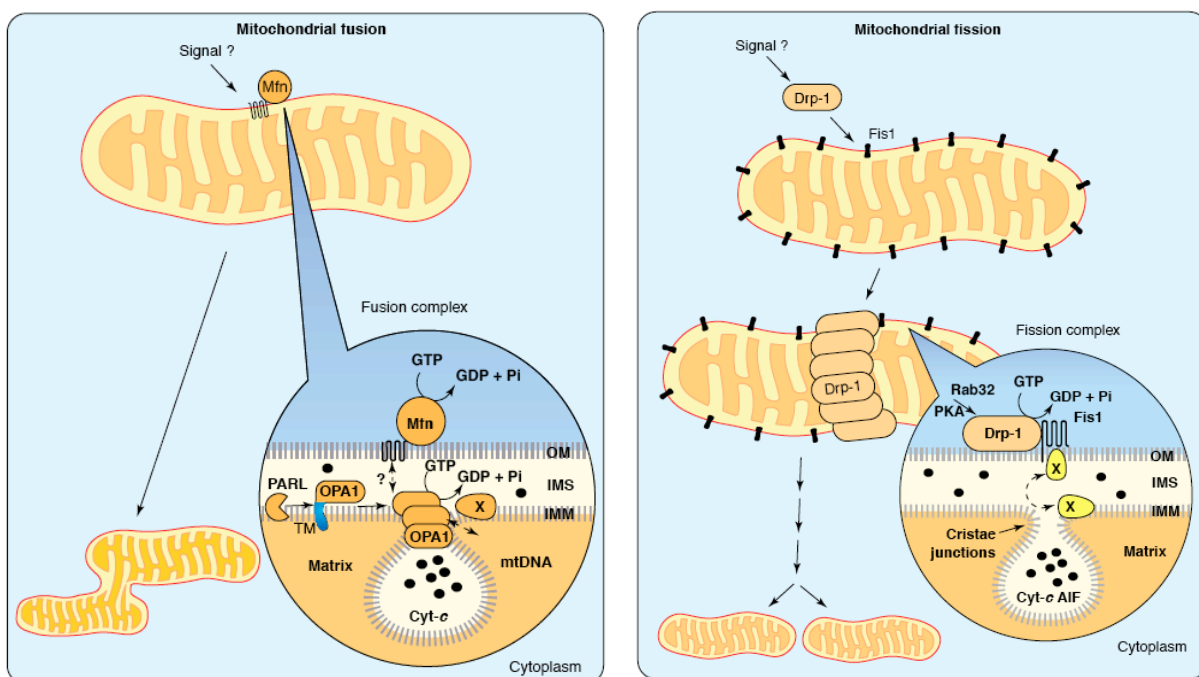
Figure 3-27. Schematic model of apoptotic mitochondrial fission. Ca2? signals from the ER or other apoptotic stimuli trigger the translocation of Drp-1 and Bax to the mitochondrial outer membrane. Bax interacts with endophilin-like molecules that contain an SH3 domain, such as Bif-1. Bax, endophilins and Drp-1 are postulated to form a large complex at sites of mitochondrial fission. Drp-1 may convey a conformational change to Bax and associated molecules, thereby facilitating an increase in outer membrane permeability and translocation of cytochrome c across the cytosol.
Inactivation of PARL or OPA-1, a factor involved in mitochondrial fusion, may lead to the opening of cristae junctions and the liberation of apoptogenic factors like cytochrome c or AIF. This process may lead to disintegration of mitochondria and their removal by autophagy. Abbreviations: cyt-c, cytochrome c; TM, transmembrane region. Picture taken from (Bossy-Wetzel et al 2003).

Chromosomal distribution of the promoters in SN and VTA neurons

The genomic distribution of nanoCAGE tag clusters along chromosomes can be visualized either by plotting the tpm density of A9 versus A10 clusters sorted by chromosome using R programming (scatter plot) or in a more visual friendly way using directly the wig tracks loaded on the Genome Browser. This allows the display of both sense and antisense TSS clusters. They are represented by vertical bars in correspondence of the TSS +1 position, having a positive value (tpm) for the tags on the (+) strand and negative value for tags on the (-) strand as shown in Figure 3-28/29 for some chromosomes. In general, there is a good overlap in the peaks of the tag cluters among A9 and A10 libraries reflecting the expected large similarity among those two cell populations. However, at a more detailed visual inspection, peaks are clearly differentially expressed among libraries as can be noted in the case of chromosomes 15 and 18, as shown in figure 3-28.

Therefore, A10 and A9 nanoCAGE tracks were manually screened selecting candidate genes for further evaluation.
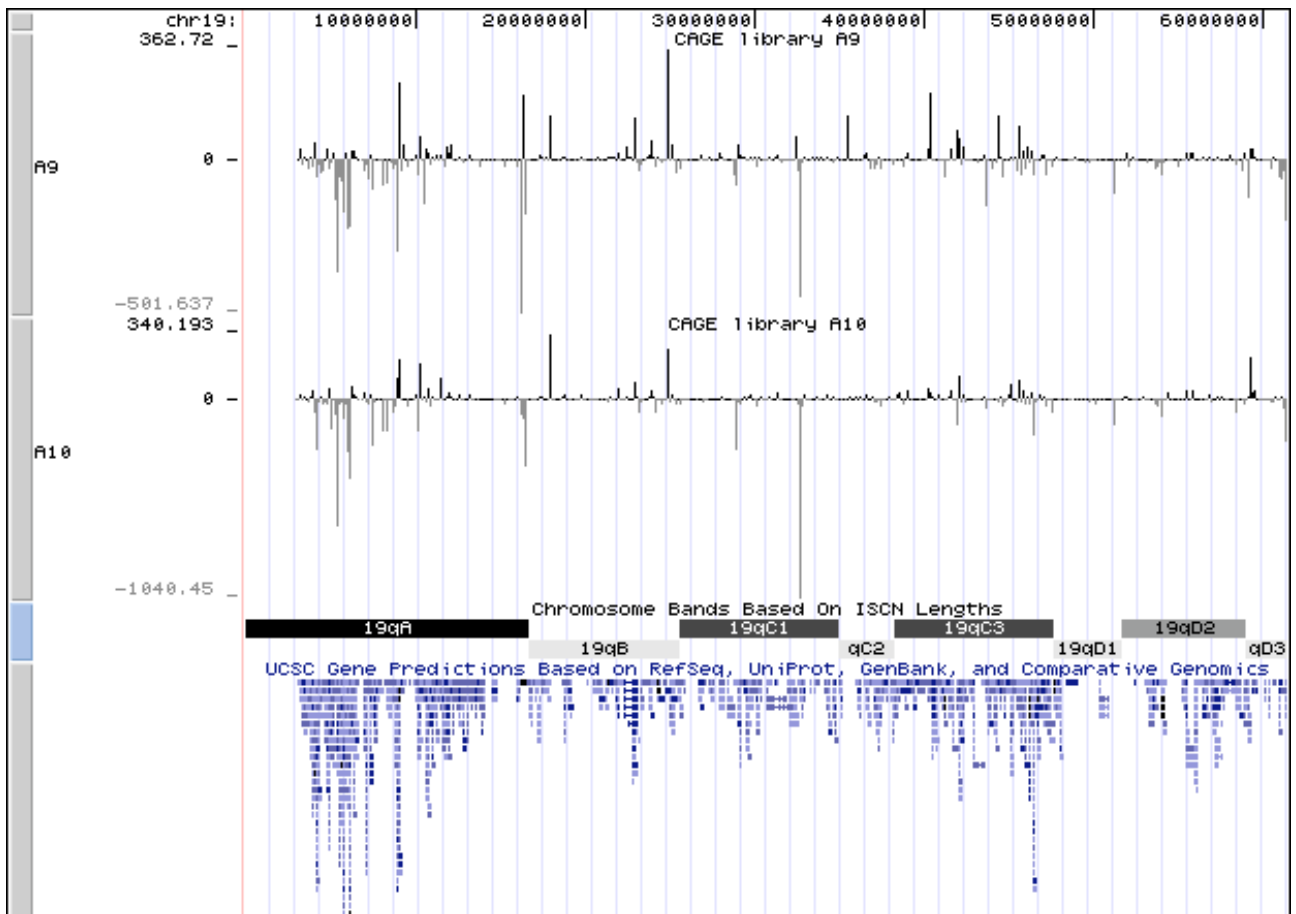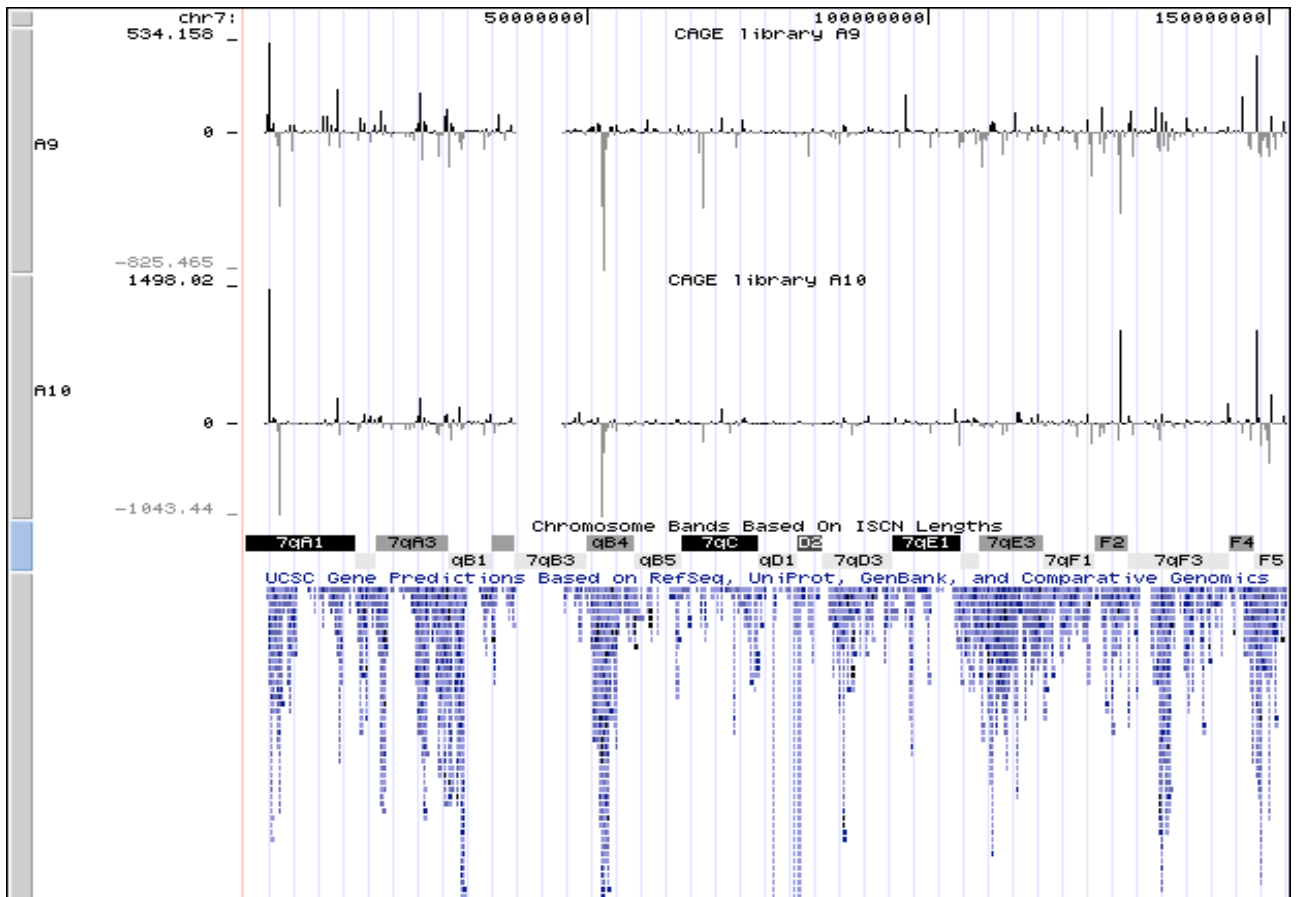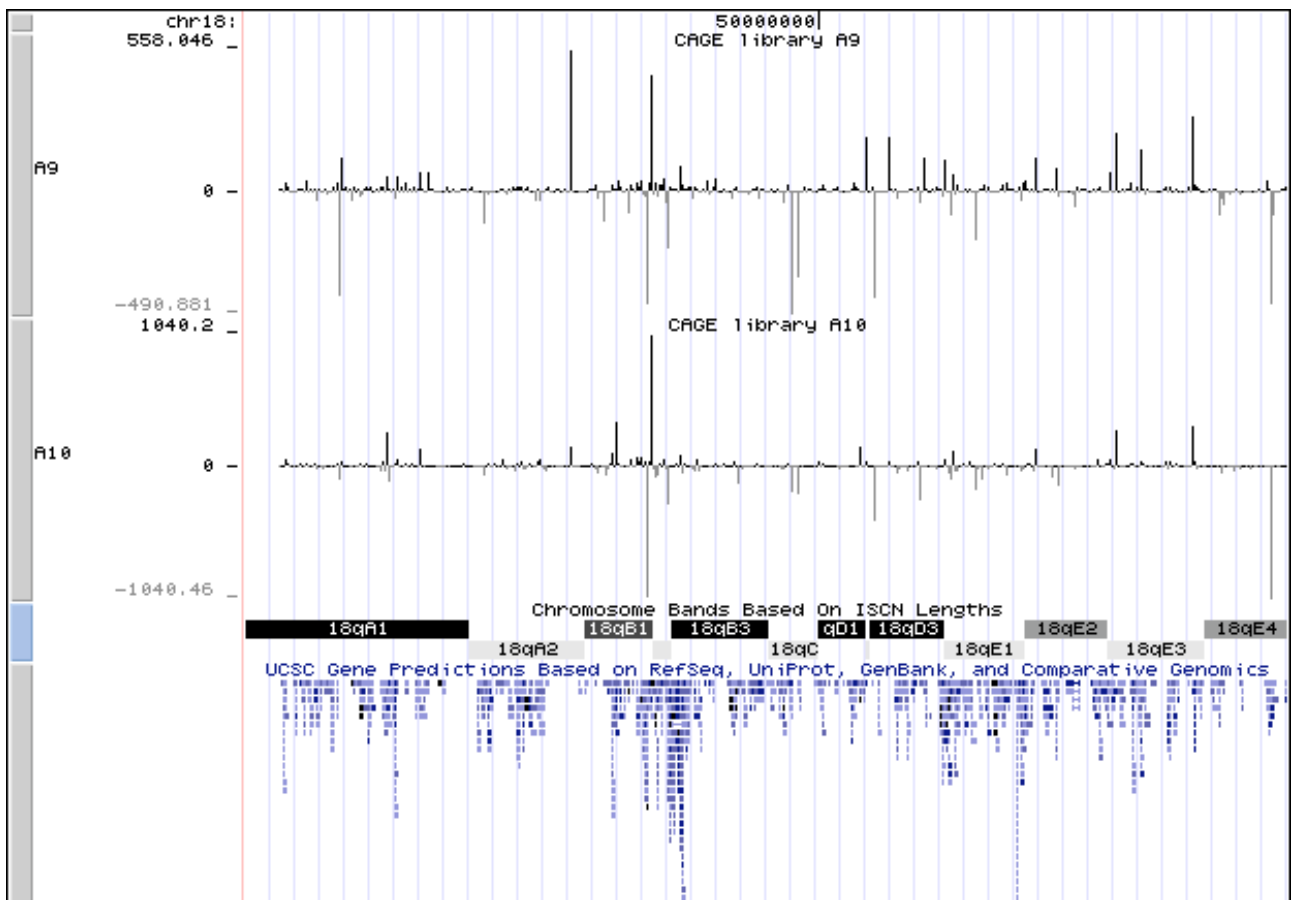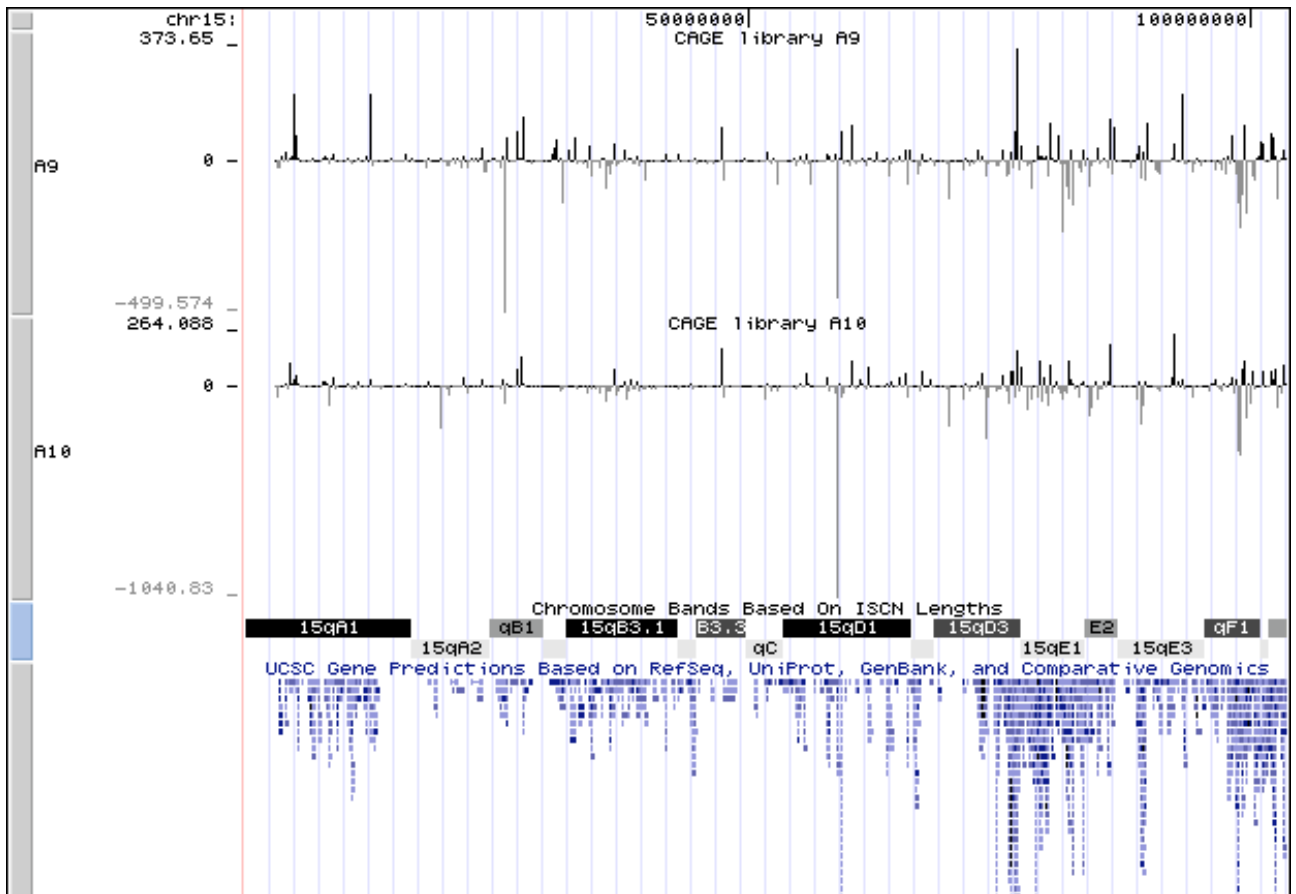
Figure 3-28. nanoCAGE tracks loaded on UCSC Genome Browser for chromosomes 7, 19, 15, 18.
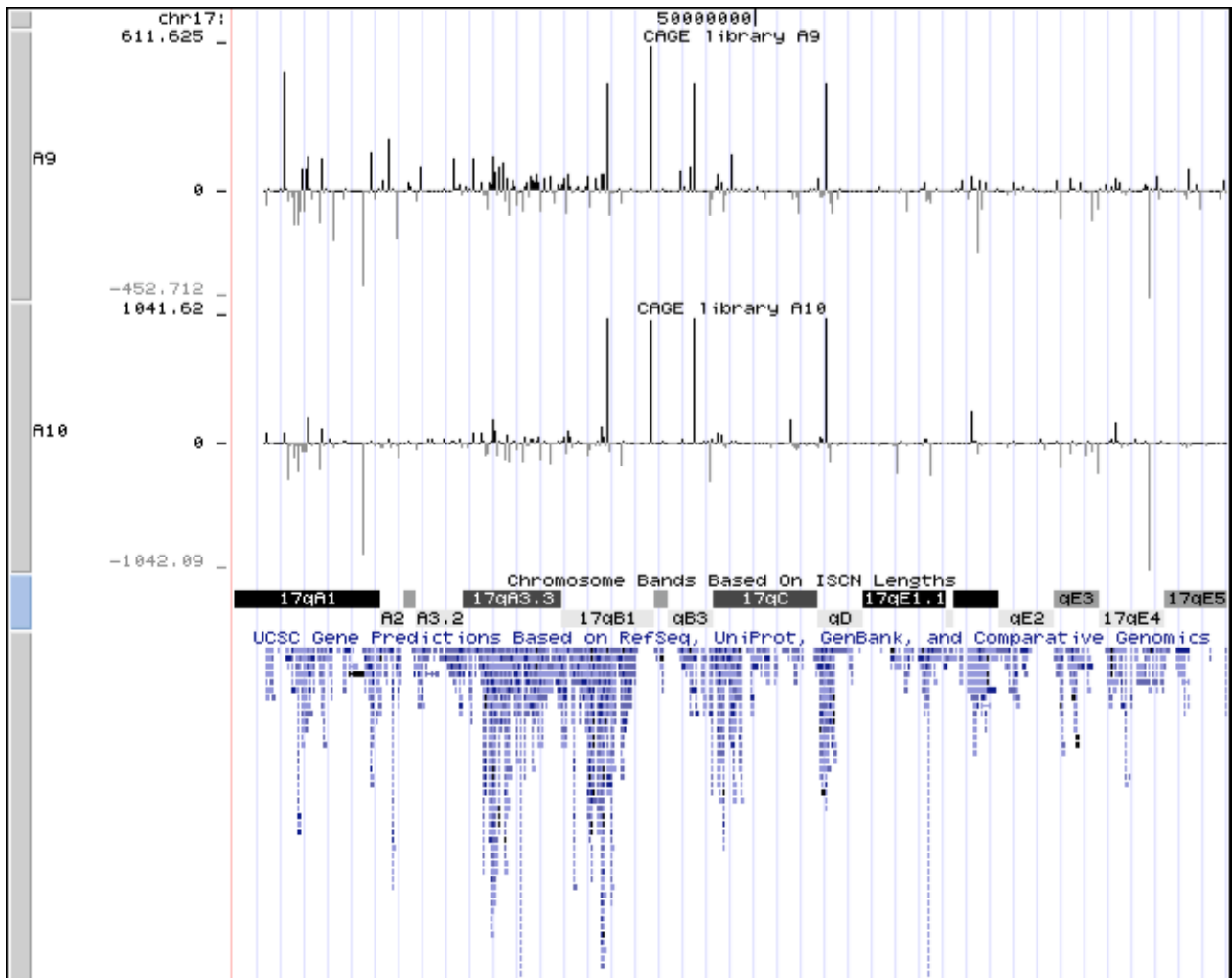
Figure 3-29. nanoCAGE tracks loaded on UCSC Genome Browser for chromosome 17

Use of alternative promoters of biologically relevant protein coding genes in SN and VTA neurons

While the analyzing differentially expressed nanoCAGE clusters associated to protein coding genes in A9 and A10 cells, it has been clear that some genes strongly implicated in regulating crucial aspects of dopaminergic cell's physiology are associated to alternative TSSs. Therefore, they may generate cell-type specific transcripts. Here we describe four cases: α-synuclein, DAT, Vmat2 and Comt.

Alpha-synuclein (Snca): Differential usage of alternative promoters potentially linked to different splicing patterns and post-transcriptional modifications of the protein

By means of nanoCAGE technology, we found that in mouse midbrain, the α-syn gene is transcribed by two main alternative TSSes, each associated to a specific type of neurons. As shown in figure 3-30, in A9 neurons transcription starts in correspondence of a more distal TSS, which is associated to the RefSeq mRNA (NM_009221), whereas in A10 neurons, transcription starts from a more proximal TSS associated to a different annotated RefSeq transcript (NM_001042451).
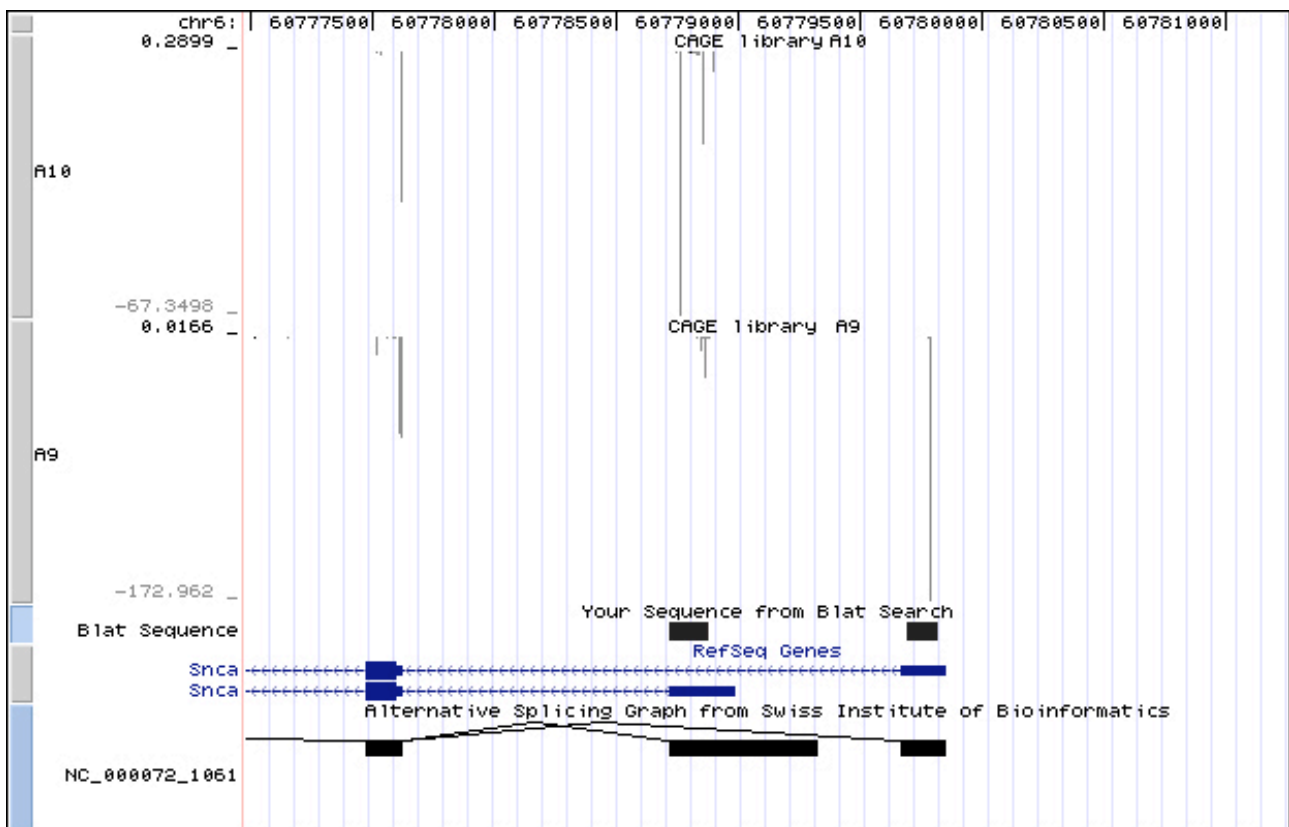
Figure 3-30. nanoCAGE tracks for A10 and A9 libraries loaded on UCSC Genome Browser displaying the tpm values of the tags clusters in correspondence of the alternative TSS of alpha-synuclein gene. Blat sequences represent the 5`RACE sequences from 2 independent clones, aligned on the mouse genome using Blat.

Therefore, RT-PCR and 5`RACE amplifications have been carried out (figure 3-31) using exon spanning primers specific for each of the two transcript isoforms (see section 2 – Materials and Methods).

We thus velidated the existence of these two transcript variants in the mesencephalon. We are currently performing real time PCR on A9 and A10 neurons to confirm their differential expression. RACE data prove that NanoCAGE technology select for true 5' ends of transcripts.
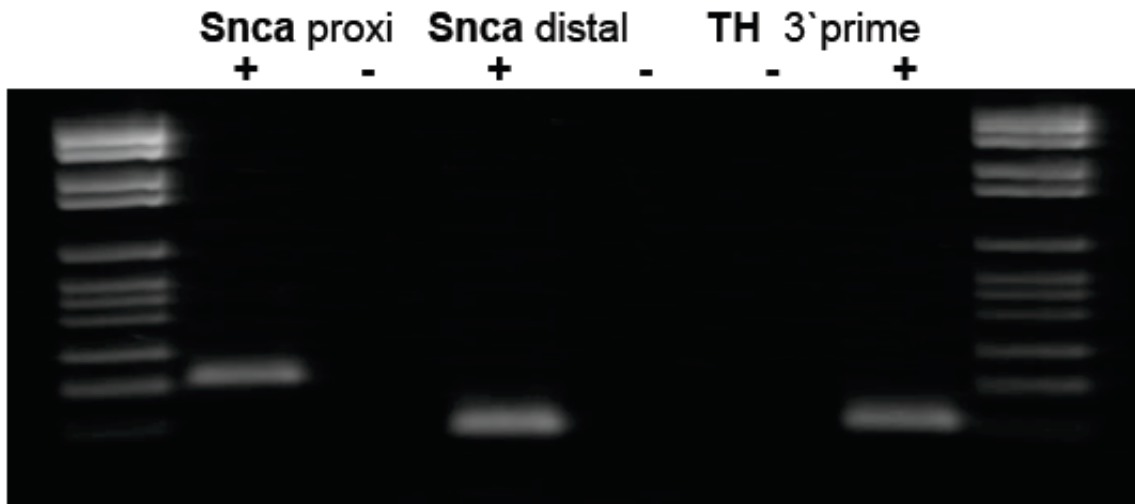
Figure 3-31. RT-PCR amplification from mouse ventral midbrain total RNA, of Snca cDNAs corresponding to the transcripts generated by alternative TSSes: distal TSS (mostly represented in A9 neurons) associated to the transcript (NM_009221) gave an amplicon of the expected size (151bp). Snca proxi represents an amplicon of the expected size (240 bp) corresponding to the transcript (NM_001042451) associated to proximal TSS (mainly used by A10 neurons). TH was used as a positive control.
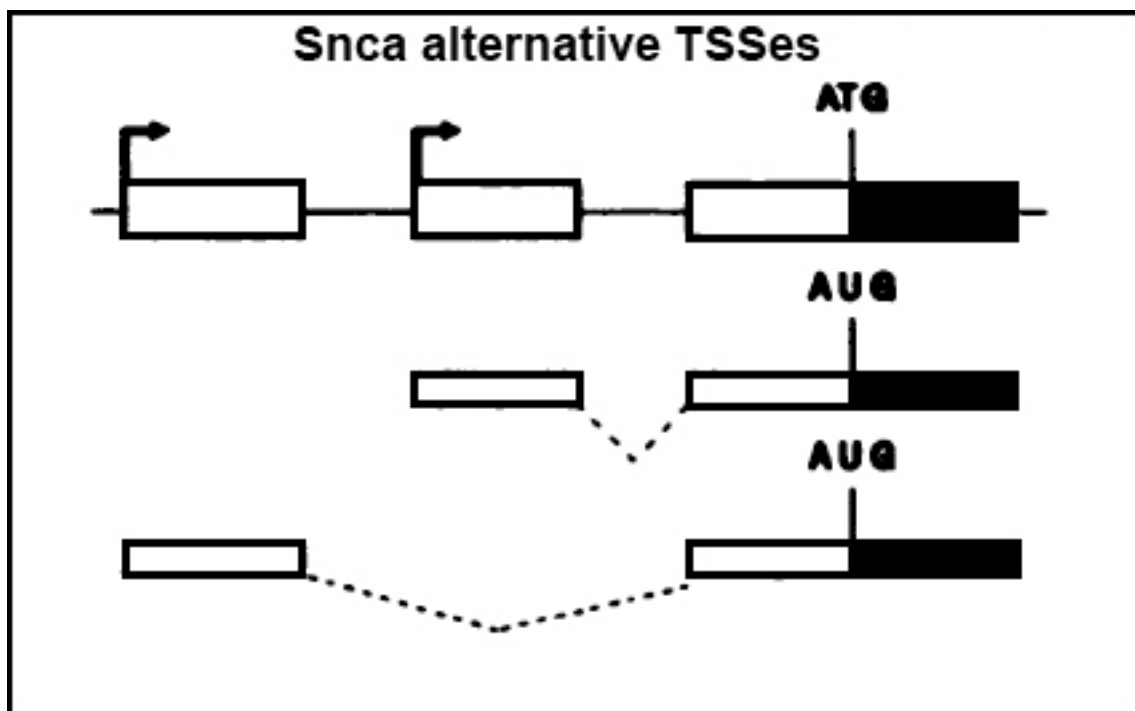
Figure 3-32. Scheme of the alternative TSSs for Snca gene. There is a difference in the splicing pattern of 5`UTR without changing the CDS. Exons are depicted as boxes. Coding regions are shaded to show origin of protein isoforms. Dotted lines connecting exons represent splicing patterns.

The two alternative promoters produced transcripts having different 5`UTR but with no change in

the coding sequence (CDS) of the respective proteins. The two alternative 5` untranslated regions were separated by approxymately 0.7 kbp distance and they differed in their predicted structures, having different fold energy (-61.30 kcal/mol for A9; and -123.40 Kcal/mol for A10) as it can be seen in figure 3-33.
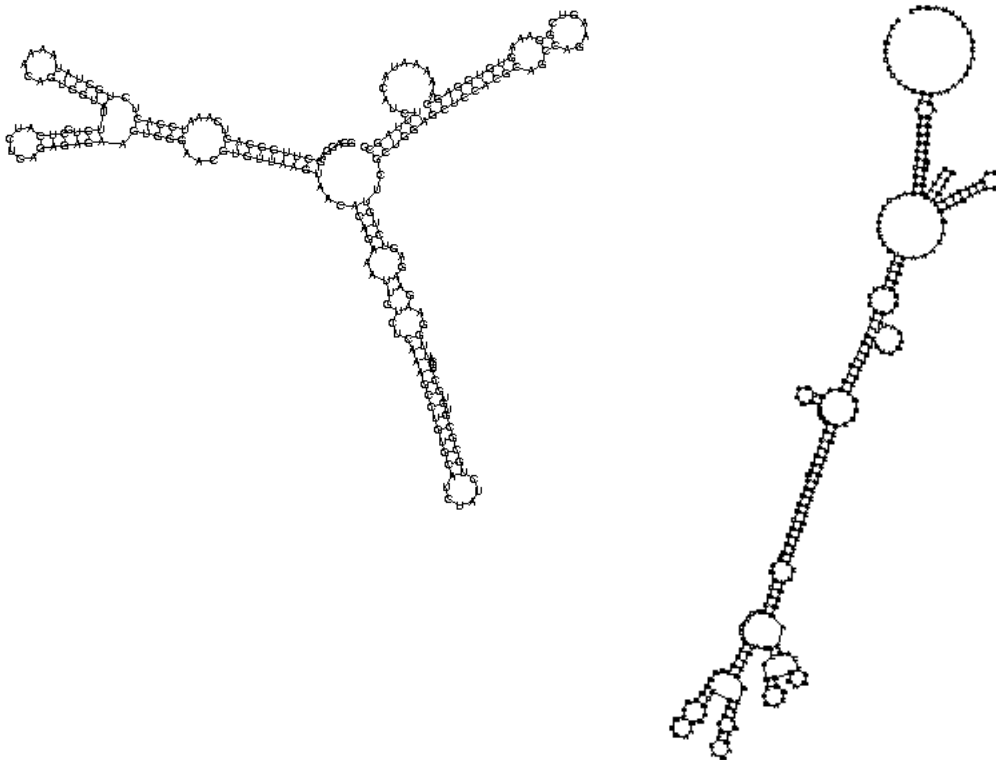
Figure 3-33. Predicted structures of the 5`UTR regions specifically associated to the A9-A10 alternative transcripts as calculated by RNA Vienna package. A9 associated transcript (NM_009221) has a more branshed but less stable 5`utr (on the left) whereas A10 associated transcipt (NM_001042451) has more compact and stable folding (on the right).
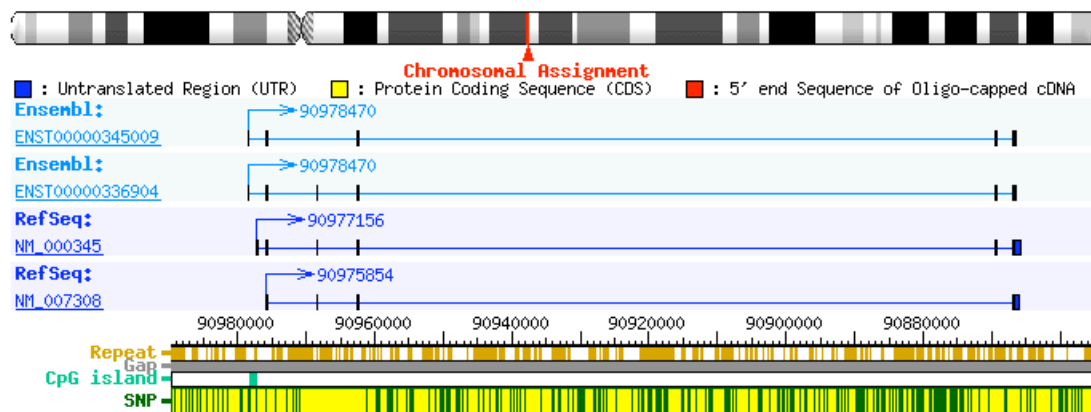
Figure 3-34. Gene structure diagram for the human SNCA locus, with Ensembl and RefSeq transcripts annotations.

Dopamine Transporter (Slc6a3/Dat): Different alternative TSSs associated to potentially different protein isoforms

Slc6a3/DAT encodes for the dopamine transporter, and it is strongly involved in regulating the normal biological functions of these neurons. We have found evidences of potential alternative TSS usage in A9 and A10 neurons. Unlike α-syn, which is ubiquitously expressed in the brain, this gene is classically considered a marker for DA-ergic cells.

The nanoCAGE libraries form A9 and A10 neurons reveal the existence of cell-type specific alternative TSSes associated to different transcripts (Figure 3-35). The main TSS in A10 map on the 4th exon at position (chr13: 73,682,261 tag seq: **GG**AACAGCCCCAACTGCTCTGATGCA), giving rise to a transcript shorter than the canonical annotated RefSeq (NM 010020), whereas in A9 the main TSS is found in correspondence of the canonical start of transcription that has been reported in other studies (Bannon et al 2001). Interestingly, the two alternative TSSs have a different initiator-like dinucleotide sequences (suboptimal GG in the case of A10 TSS, canonical pyrimidine-purine TG in the case of the highly represented TSS of A9).

Importantly, both A9- and A10-associated TSSs were confirmed by 5`RACE analysis as it can be seen in figures 3-35, 3-45. This proves again that nanoCAGE technology select for true 5' ends of transcripts.

Interestingly, as shown before in section 1 (Page 114-115), at least in human there are evidences of multiple transcripts at the SLC6A3 locus, many of which have a coding potential as demonstrated by the presence of in frame methionine residues.

Figure 3-35. nanoCAGE tracks of A10 and A9 libraries loaded on UCSC Genome Browser displaying the tpm values of the tags clusters in correspondence of the alternative TSS of Slc6a3 gene. Samples sequences over the RefSeq gene represent the 5`RACE sequences from 2 independent clones, aligned on the mouse genome using Blat.
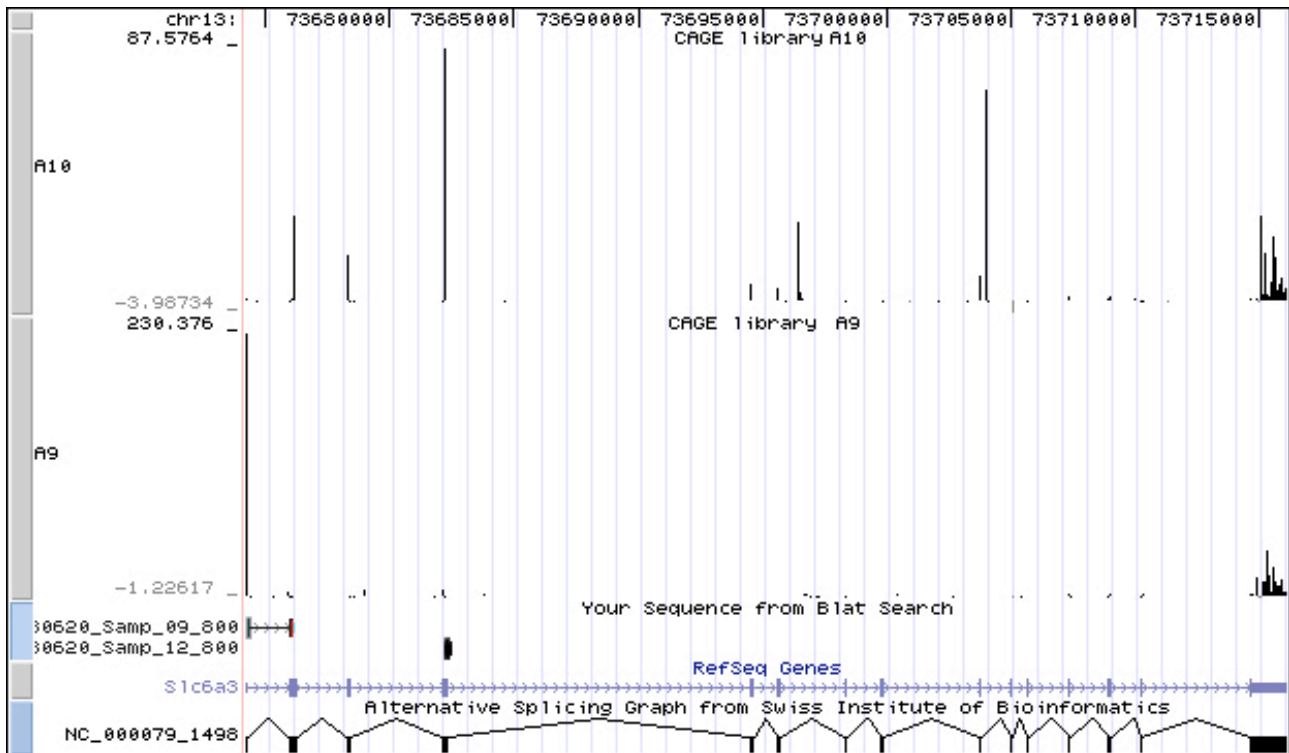
Vescicular Monoamine Transporter (Slc18a2): an alternative intronic TSS is mainly used by A10 neurons

Slc18a2/VMAT2 encodes for the vescicular monoamine transporter, a protein involved in regulating the monoaminergic vescicular storage. It has been found to use alternative TSSes in the two DA cell populations of A9 and A10.

The analysis of the nanoCAGE tracks for Slc18a2/Vmat2 murine gene revealed that two main TSSes are specifically used by the A9 and A10 neurons. The TSS used in A10 neurons lies within the intron 3, an exceptionally long intron spanning more than 10 kbps, whereas in A9 neurons the more 5`distal TSS is the canonical one, reported in literature and associated to the 5`end of the respective RefSeq transcript (NM_172523). Interestingly, as in the case for DAT, the two alternative TSSes seems to have a different initiator-like dinucleotide sequences (a canonical TG in the case of A10 TSS, tag seq: TGCTTCTTTCTGGTTCAGGGTCTGTGC; and CA in the case of the highly represented TSS of A9). Both A9- and A10-associated TSSes were confirmed by 5`RACE and RT-PCR analysis as it can be seen in figure 3-36.

212
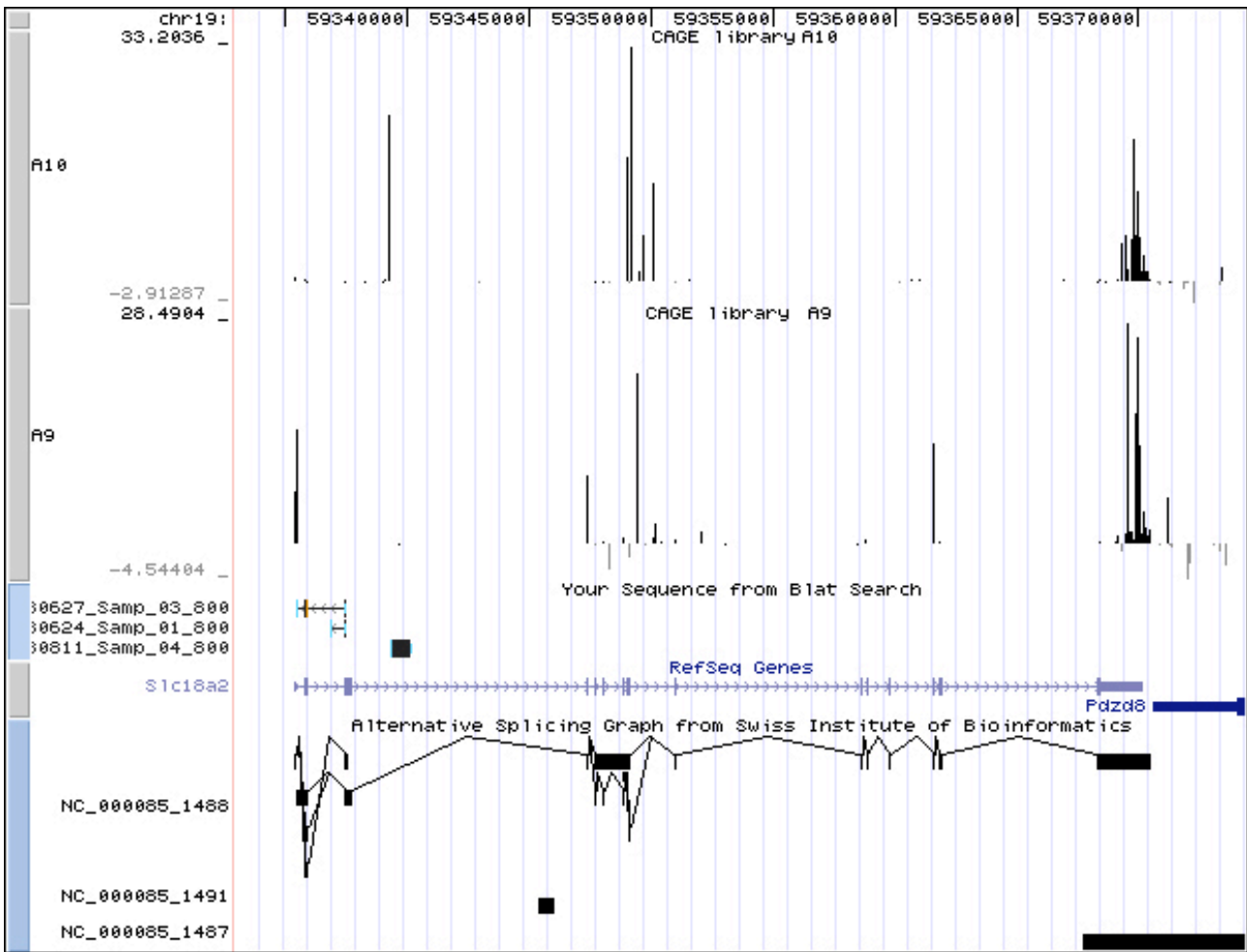
Figure 3-36. nanoCAGE tracks of A10 and A9 libraries loaded on UCSC Genome Browser displaying the tpm values of the tags clusters in correspondence of the alternative TSS of Slc18a2 gene. Sample sequences over the RefSeq gene represent the 5`RACE sequences from 3 independent clones, aligned on the mouse genome using Blat. An additional 5`RACE product has been cloned, corresponding to the TSS for other previously isolated cDNA clones (Jassen et al 2005)

At least for the human gene another isoform of the VMAT2 transcripts have been reported as shown by the informations retrieved from the AceView database (Figure 3-37).
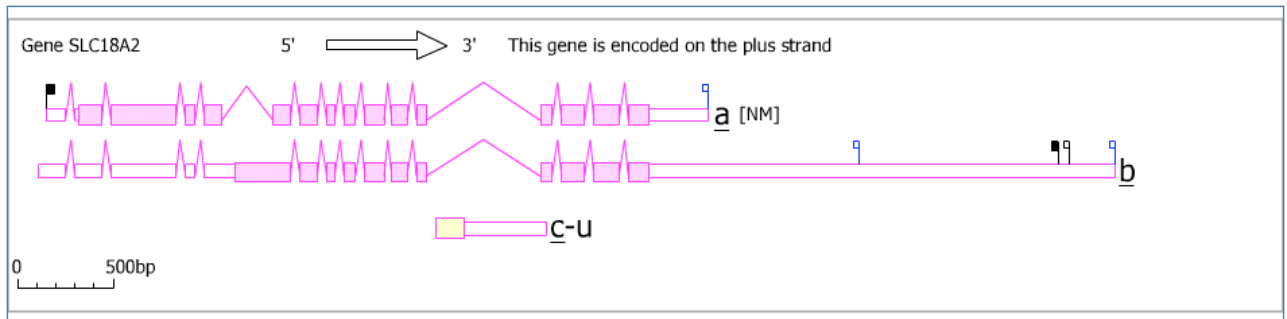
Figure 3-37. A scheme of the possibile alternative transcripts and their corresponding proteins associated to the human Slc18a2 locus as emerges from cDNA and EST data collected from many tissues and summarized by the AceView gene prediction database. At least 3 different isoforms are reported for VMAT2; the isoform a (12 accessions from substantia nigra, brainstem, colon, embryo) isoform b (is the most represented with 46 accessions from pancreas, head neck, placenta, melanocytes, fetal heart, pregnant uterus) and isoform c (4 accessions from liver). Picture taken from Aceview database.

The 5`RACE amplification have been done using exonic gene specific primers downstream of each

TSS (distal or proximal) of the two transcript isoforms (see Materials and Methods) (figure 3-38).

We are currently validate the cell-type expression by real time PCR on LCM-purified cells.
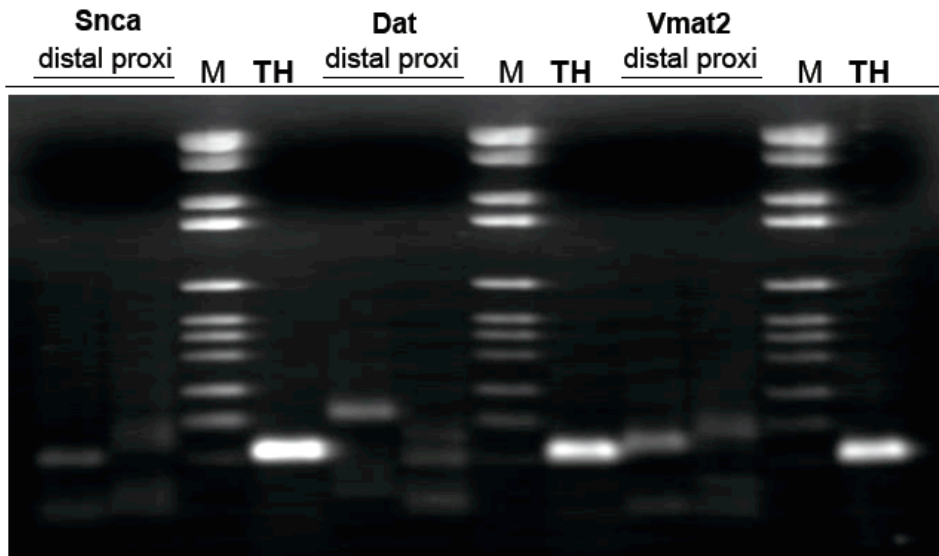
TH is a Control RT-PCR on the same cDNA

Figure 3-38. 5`RACE amplification from mouse ventral midbrain total RNA, of specific cDNAs corresponding to the transcripts generated by alternative TSSes: Snca distal (mostly represented in A9 neurons) associated to the transcript (NM_009221); Snca proxi (mostly represented in A10) associated to the transcript (NM_001042451); Dat distal (mainly represented in A9) starting in 1st exon; Dat proxi (mostly present in A10) starting in the 4th exon; Vmat2 distal (mainly used by A9 neurons) from the 1st exon; Vmat2 proxi (A10) starting within 3rd intron; TH represents an RT-PCR amplicon obtained from the same cDNA, used as a positive control.
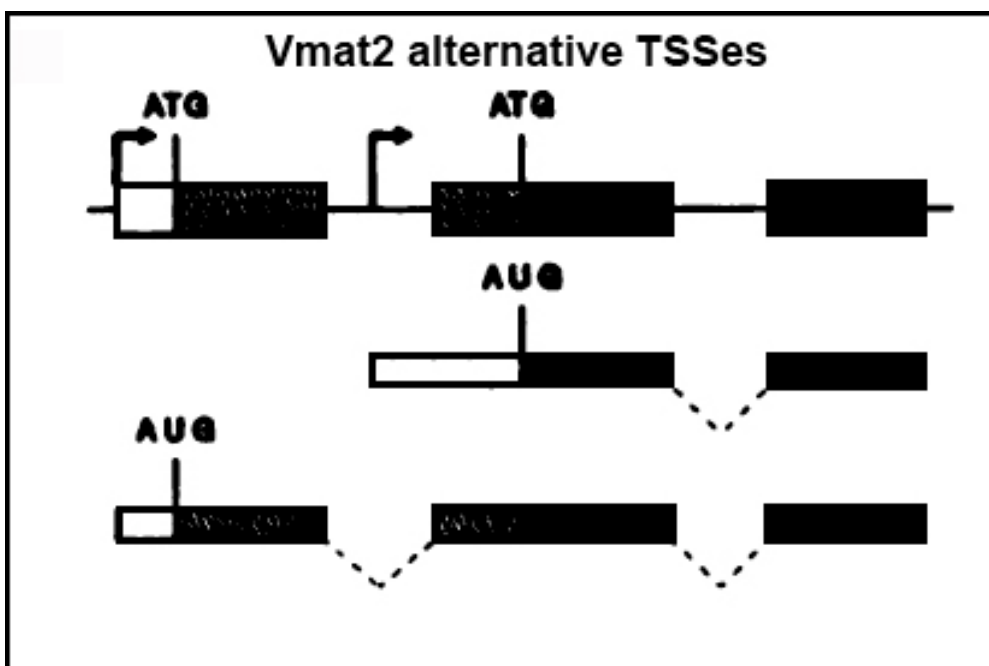
Figure 3-39. Scheme of the alternative TSSes for Slc18a32 gene. There is a difference in the splicing pattern of 5`UTR and there is a drastical change of the CDS (to simplify the scheme the actual number of exons is not reported). Exons are depicted as boxes. Coding regions are shaded to show origin of protein isoforms. Dotted line are introns.

Catechol-O-methyltransferase (Comt): alternative promoter usage associated to MB-Comt and S-Comt in A9 and A10 dopaminergic neurons

Another gene, Comt, coding for the enzyme catechol-O-methyltransferase has been found to have a differential expression from alternative promoters in A9 and A10 DA neurons.

From the nanoCAGE libraries, the two transcripts corresponding to the S-COMT and MB-COMT are differentially expressed among A10 and A9 neurons.
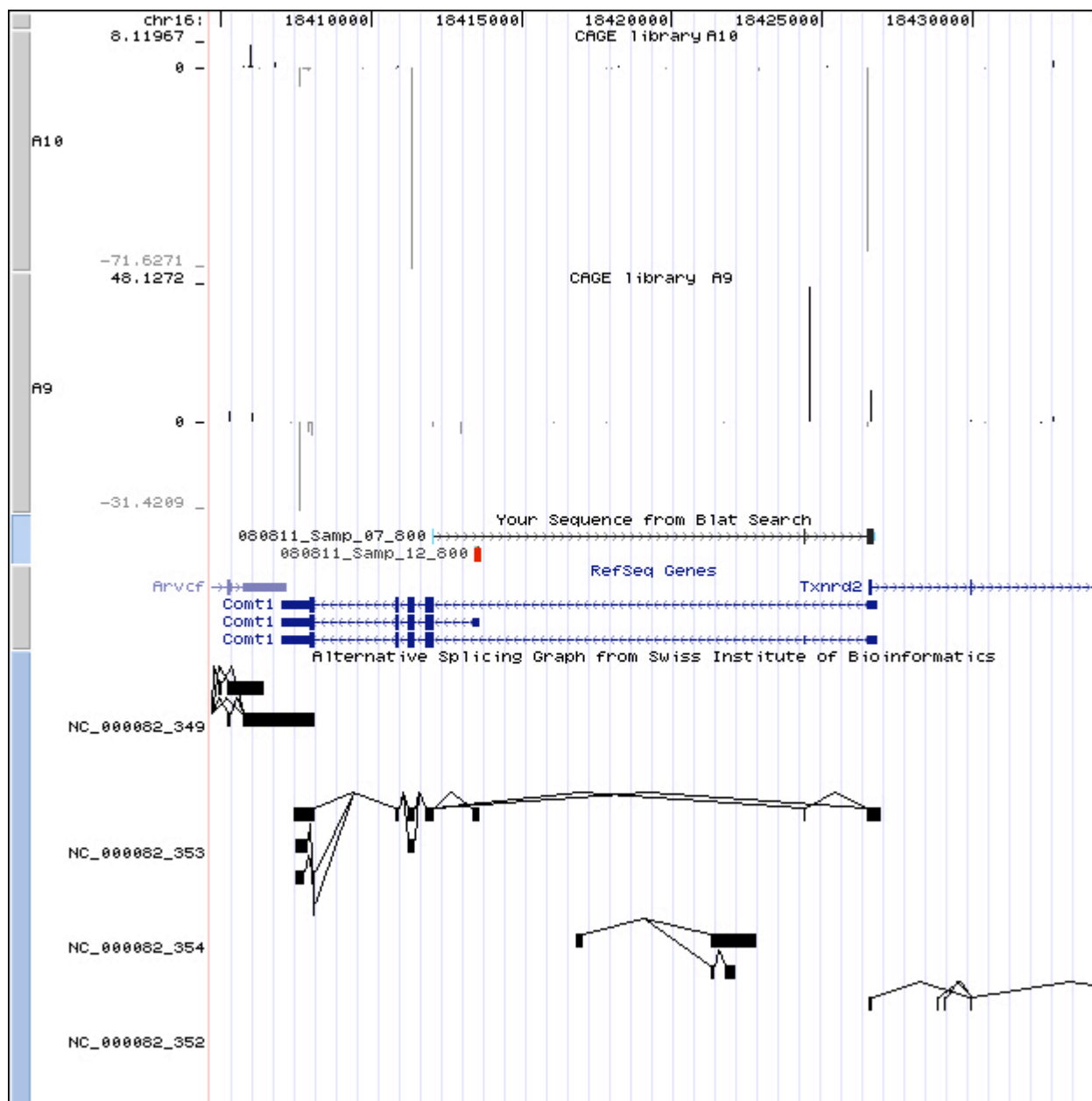
Figure 3-40. nanoCAGE tracks of A10 and A9 libraries loaded on UCSC Genome Browser displaying the tpm values of the tags clusters in correspondence of the alternative TSS of Comt gene. Sample sequences over the RefSeq gene represent the 5`RACE sequences from 2 independent clones, aligned on the mouse genome using Blat .

215
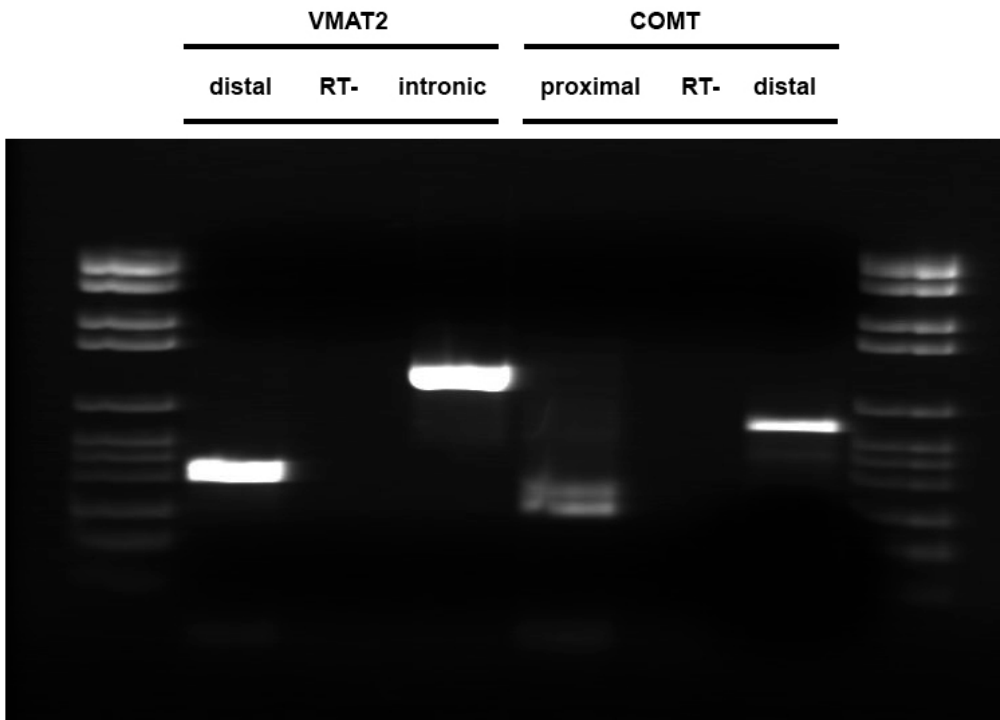
Figure 3-41. RT-PCR analysis of the alternative transcript isoforms associated to the Vmat2 and Comt mouse genes. In the first lane an amplicon of the expected size (199 bp) was obtained for Vmat2 isoform starting from a distal TSS (Vmat2 distal); an amplicon of the expected size (391 bp) was also obtained for the Vmat2 isoform starting from the intronic TSS (Vmat2 intronic) with RT- giving no signal for the same intronic isoform; Comt isoform transcribed from the internal TSS (Comt proximal) gave an amplicon of expected size (179 bp); Comt long isoform transcribe from the distal TSS (Comt distal) gave an amplicon of approximately 291 bp.
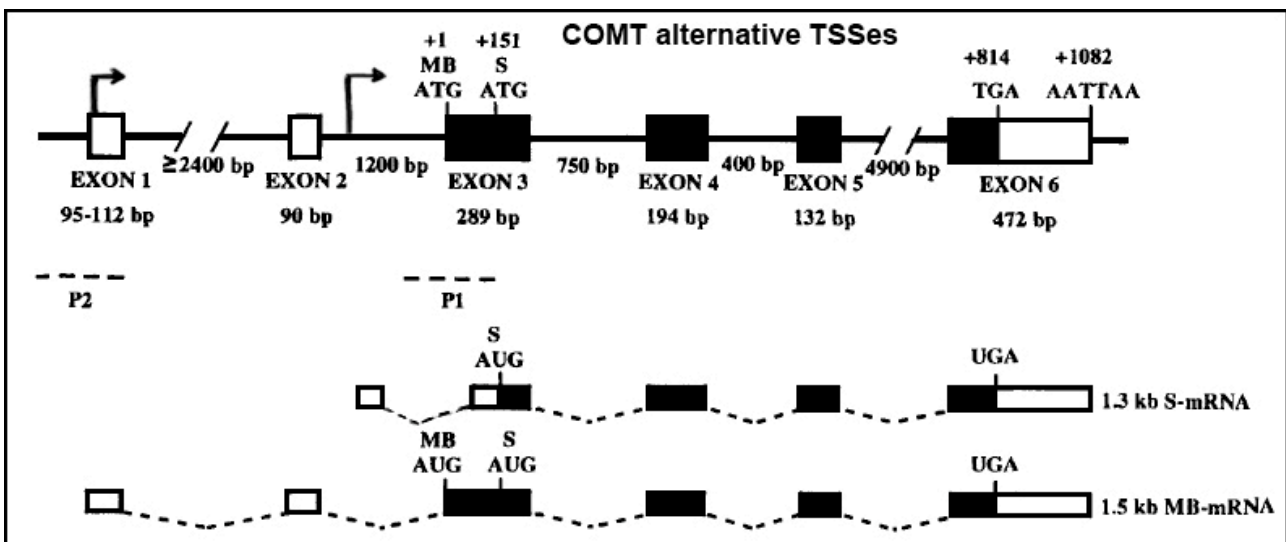
Figure 3-42. Scheme of the alternative TSSs for Comt gene. There is a difference in the splicing pattern of 5`UTR but and a corresponding change of the CDS, with the longer ORF coding for a membrane-binding protein, and the shorter ORF coding for a soluble protein. Exons are depicted as boxes. Coding regions are shaded to show origin of protein isoforms. Dotted lines connecting exons represent splicing patterns.

In A10 neurons both isoforms are expressed at relatively high level, whereas in A9 neurons only the S-COMT form appears to be expressed at low levels. Interestingly, an antisense transcript is present at the distal TSS associated to the MB-COMT isoform. A 5`RACE analysis has confirmed the existence of both the TSS associated to the alternative promoters, as shown in figures 3-40 / 3-41.

The non coding RNAs landscape in SN and VTA neurons as emerging from nanoCAGE and deep sequencing

Relative abundance and different families of non coding RNAs

The relative abundance of the non coding transcripts over the total count numbers has been calculated and plotted for each class of transcripts, ncRNAs (F3 conservative vote set, as reported before), RefSeq mRNA, miRNAs (mirBASE), and for the two different clustering methods (Parametric clustering - PC, and Tag proximity clustering - TC), (Figures 3-43, 3-44)
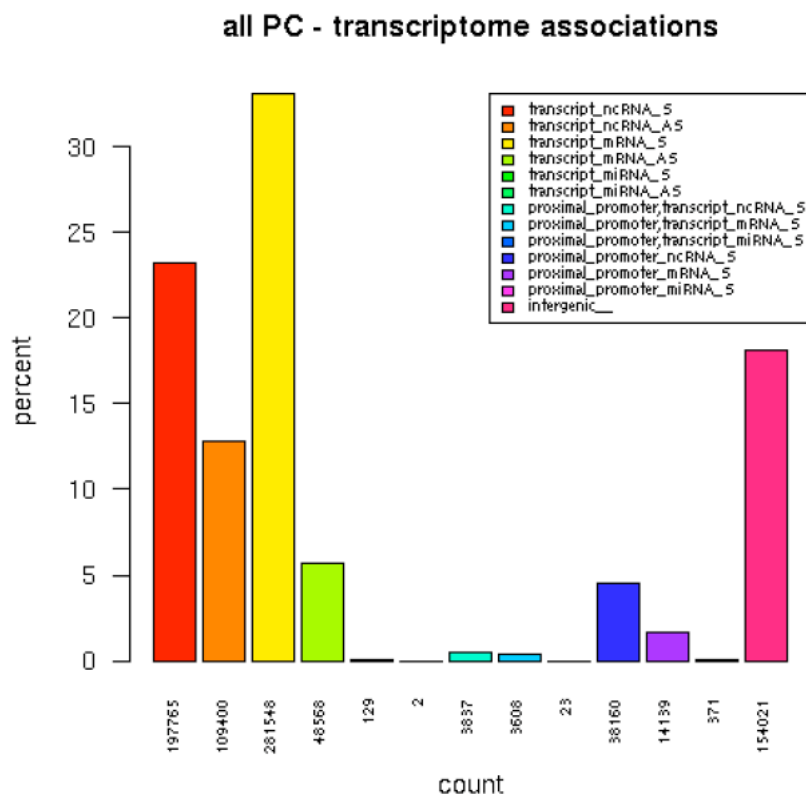
Figure 3-43. It appears that ncRNAs account for about 36% of the transcriptome, when considering all the PC, clusters, irregardless of their stability score or the uniquess location in the genome. Whether coding RNAs account for about 38,8% of the transcriptome.
nanoCAGE Tags matching intergenic locations in the genome are also quite an abundant class, accounting for about 18% of the total counts. However most of the intergenic tags are less representative and statistically much less relevant, due their low stability score.

217

Here you can find the association of tags clusters obtained with different clustering methods with current transcriptome annotations:

Figure 3-44
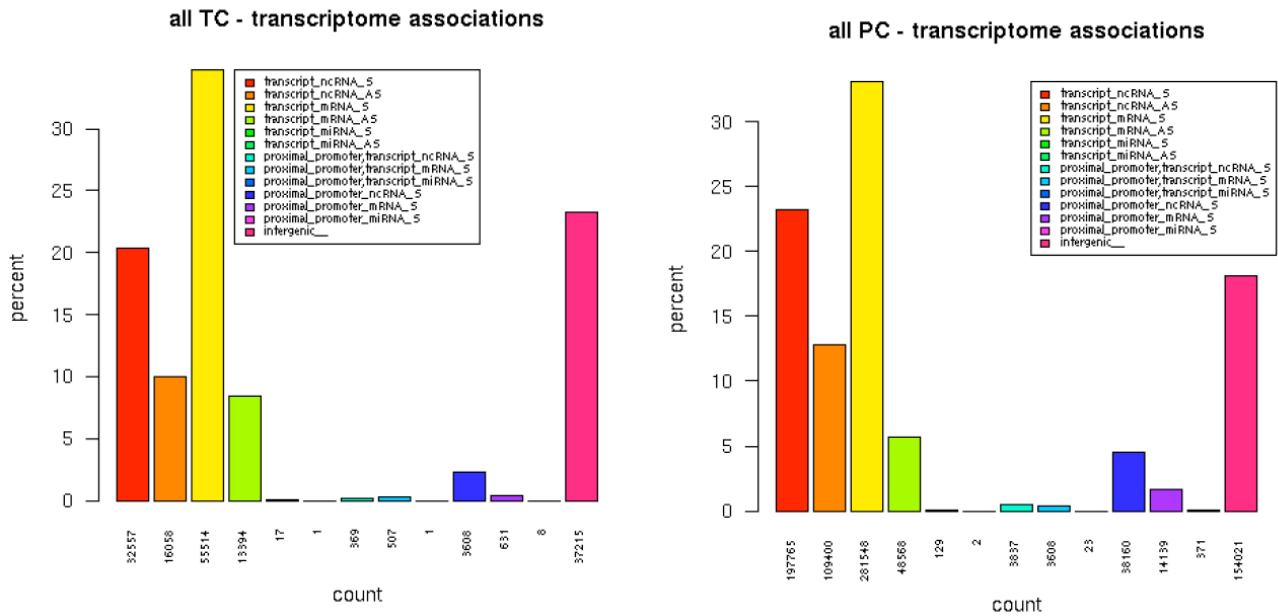
The distribution among different classes of transcripts gave similar results for the two different clustering methods (PC or TC). However, when we considered only tag clusters that have been stable and non ambigously positioned in the genome, an higher number of cluster associated to known transcripts was visible.
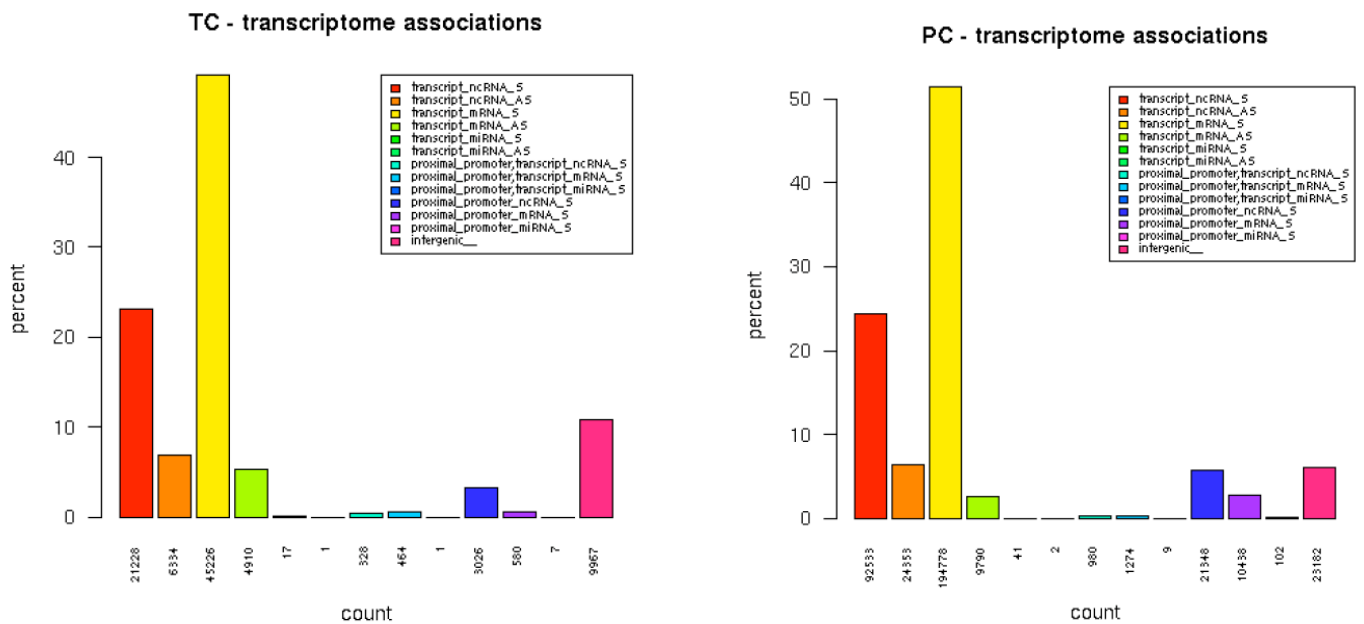
Figure 3-45

Differentially expressed microRNA transcripts in SN and VTA neurons

A list of putative differentially expressed miRNA transcripts has been compiled, searching for tag clusters lying in an interval of [-500..-1] bp from the annotated miRNA precursor sequence (coordinates were retrieved from miRBASE). The tpm values were then compiled according to two methods:, 1. counting the tags matching unambiguously in only one position, 2. considering the tags that map in more than one position in the genome. You can find the list of the most significantly expressed miRNA in A9 and A10 neurons in a decreasing order.

Figure 3-46.
Here are listed only 17 entries, which have shown the most different values.

It can be seen that some of these miRNA seem to be expressed at a significantly different level, like in the case of mir-677 and mir-671.

However, it has to be noted that the interval of 500bp upstream of the TSS of the precursor is likely to not comprise the promoter of the

name-A9_single_multi_tpm-A10_si			name-A9_single_only_tpm-A10_sing		
mmu-mir-682	1738.48	957.048	mmu-mir-682	3753.21	1896.77
mmu-mir-692-2	477.301	101.697	mmu-mir-677	258.794	6.429
mmu-mir-677	106.663	2.199	mmu-mir-678	101.064	226.141
mmu-mir-685	93.529	49.584	mmu-mir-685	210.017	122.143
mmu-mir-678	41.617	77.369	mmu-mir-219-2	75.662	10.419
mmu-mir-703	8.29	7.729	mmu-mir-673	48.918	74.041
mmu-mir-684-1	13.561	3.331	mmu-mir-671	71.328	1.108
mmu-mir-219-2	31.161	3.564	mmu-mir-770	62.289	60.077
mmu-mir-671	29.373	0.379	mmu-mir-207	5.457	44.563
mmu-mir-7a-1	0	0	mmu-mir-708	0	0
mmu-mir-770	25.883	20.55	mmu-mir-126	0	39.021
mmu-mir-673	20.139	25.323	mmu-mir-423	12.044	9.312
mmu-mir-692-1	17.435	25.104	mmu-mir-705	4.517	0
mmu-mir-684-2	1.007	1.135	mmu-mir-200b	0	0
mmu-mir-207	3.484	16.534	mmu-mir-429	0	0
mmu-mir-708	0	0	mmu-mir-141	0	0
mmu-mir-423	4.959	3.186	mmu-mir-697	0	11.085

primary transcript of the microRNA, as it has been demonstrated that in the majority of the cases this is located in a range of 2-10 kilobases, from the precursor microRNA sequence. Since not so much data are currently available in databases that can be used for a more accurate bioinformatics analysis, these results should be considered only indicative. A more rigorous analysis of experimentally determined list of promoters should be reconsidered to refine the list.

Sense-Antisense pairs transcription

To uncover potentially interesting antisense transcripts, all the sense-antisense pairs where a non coding antisense RNA was overlapping in the first intron/exon of a protein coding transcript were selected, using a perl script. Then a further selection was based on the difference in the normalized expression level (tpm) among A9 and A10.

```
ncrna_name  ncrna_strand  ncrna_A9_tpm  ncrna_A10_tpm  refseq  refseq_proximal_A9_tpm  refseq_distal_A9_tpm  refseq_proximal_A10_tpm  refseq_distal_A10_tpm  coord
B230219N05   -   56.01  0.16  NM_016663      0      54.32   0       46.42   chr7:51634257..51655399   (Syt3/Synaptotagmin III)
A730037D09   -   56.01  0.16  NM_016663      0      54.32   0       46.42   chr7:51635757..51655399   (4-hydroxyphenylpyruvic acid dioxygenase/ Hpd) Defects cause of tyrosinemia type III
D930046A11   -   32.8   2.93  NM_027297      0      0.91    0       131.59  chr4:62060017..62104081
B230345005   +   27.35  0.12  NM_025842   92.55  58.36  192.88  35.87   chr15:76452517..76458690  AS to Nfkbil2/I-kappa-B-related protein
K430346H17   -   22.7   2.24  NM_001037756   0      6.07    0       14.38   chr12:56917692..57113461
C630024B11   -   20.16  1.43  NM_008829      0      3.1     0       2.51    chr9:8899520..8968611     AS to Pgr/progesterone receptor
F430102K23   -   16.3   0     NM_009439      0      23.25   0       92.32   chr11:98543046..98557292  AS to Psmd3/proteasome 26S non-ATPase subunit 3
C230097F24   +   12.21  0.24  NM_207000      0      5.18    0       10.67   chr10:61201394..61247568  AS to H2afy2/H2A histone family, member Y2
G370057A12   -   11.33  0.84  NM_177350      0      1.16    0       2.03    chr9:54116385..54189584   AS to Gldn/Gliomedin (low level, not specific of A9)
C820006M04   -   11.05  0.63  NM_001038660   0      0       0       0.08    chr12:8518007..8569487    AS to Slc7a15/aromatic-preferring amino acid transporter
C630032D20   -   10.96  0.12  NM_133655   344.61 210.47  274.59  151.07  chr7:150207689..150253829 AS upstream to CD81
C920004C22   -   6.95   0.16  NM_022000      0     182.04   0      328.58   chr2:174106738..174172243  AS to Gnas isoform e
C920004C22   -   6.95   0.16  NM_019690      0      6.01     0      122.61   chr2:174106738..174163585  AS to Gnas isoform c
A230076B10   +   6.9    0.08  NM_178793      0      88.2     0       31.21   chr18:66216509..66479261  AS to Ccbe1/collagen and calcium binding EGF domains 1 (specific to A10)
E230006A20   +   6.88   0.08  NM_021876      0      28.28    0       12.26   chr7:97103185..97131330   AS to Eed/embryonic ectoderm development
D630010J15   +   6.1    0     NM_008732      0      0.94     0.87    5.94    chr15:100218330..100257906 AS to Slc11a2/mportant in metal transport, in particular iron
C130054G18   -   5.87   0.08  NM_057171      0      35.44    0      102.46   chr17:35271634..35284181  AS to Bat3/HLA-B-associated transcript 3, contains an Ubiquitin-like domain
A530078N24   -   5.56   0     NM_008929   0.08      1.7      0       7.28    chr14:119336020..119379121 AS to Dnajc3a/DnaJ (Hsp40) homolog, subfamily C, member 3B, Up-regulated in ER stress
B630007K22   +   5.56   0     NM_009833      0      1.15     0.4     1.23    chr15:98373641..98400014  AS to Ccnt1/cyclin T1 (not specific to A9)
9230028E05   -   5.56   0     NM_133743      0      0        0       0       chr7:25416164..25426137   AS to Lypd3/Ly6/Plaur domain containing 3, Supports cell migration.
5730407N16   -   4.9    0.24  NM_028636      0     75.19     0       7.11    chr9:56947755..56990017   AS to Man2c1/mannosidase, alpha, class 2C, member 1
A830092P13   -   4.25   0     NM_172757      0     11.02     0      14.81    chr8:90661419..90695842   AS to Heatr3/HEAT repeat containing 3
2410133F24   -   4.17   0     NM_028636      0     75.19     0       7.11    chr9:56977802..56990017
1700105P06   +   3.94   0     NM_001033139   0      4.39     0      16.14    chr19:7432179..7458040    AS to AI846148 hypothetical protein
B230205G07   -   3.88   0     NM_007965      0     31.82     0      19.37    chr12:109774610..109926725 AS to Evl/Ena-vasodilator stimulated phosphoprotein
A830011D19   -   2.94   0.16  NM_183221      0      4.75     0       2.5     chr3:38773492..38910905   AS to Fat4/FAT tumor suppressor homolog 4
C030017D09   -   2.94   0     NM_175455      0      0.08     0       4.67    chr13:93184972..93210180  AS to Ankrd34b/cytosolic phosphoprotein DP58
C230034021   -   2.94   0.16  NM_183221      0      4.75     0       2.5     chr3:38776709..38910905
C030017P06   -   2.94   0     NM_175455      0      0.08     0       4.67    chr13:93184972..93210180
5830426D21   +   2.39   0.08  NM_023190      0     59.64     0      45.32    chr14:55260997..55307853  AS to Acin1/apoptotic chromatin condensation inducer 1 (very low expressed)
C920026L21   +   2.39   0     NM_178709      0      3.31     0.08   11.3     chr9:45671773..45715557   AS to Rnf14/ring finger protein 214 and sense to Pcsk7/proprotein convertase subtilisin
4930465C05   -   2.23   0.18  NM_146414      0      0        0       0       chr19:12276849..12310346  AS to Olfr1431
A830059M14   +   2.2    0     NM_172691      0     29.2      0      22.26    chr4:43045381..43066125   AS to B230312A22Rik hypothetical protein
B930005010   -   1.93   0     NM_028769      0      9.73     0       4.35    chr19:6045800..6053710    AS to Synoviolin apoptosis inhibitor 1 E3 ubiquitin ligase in ERAD pathway
1700000E03   -   1.93   0     NM_026441      0     40.96     0      41.12    chr4:129784522..129805378 AS to PEF protein with a long N-terminal hydrophobic
G630040K20   +   1.78   0     NM_025710      0     10.97     0     130.88    chr13:30632180..30639272  low in A9,  ubiquinol-cytochrome c reductase, Rieske
4833404H03   -   1.7    0     NM_013840      0      2.55     0       0.08    chrX:20528790..20563849   AS to Uxt, ubiquitously expressed transcript
1700081J07   -   1.39   0     NM_018828      0     20.4      0       1.04    chr2:90585306..90621177   AS to Fnbp4 formin binding protein 4
D130058E05   +   1.31   0     NM_010262      0      0        0       0       chr1:91824536..91829865   AS to Gbx2
```
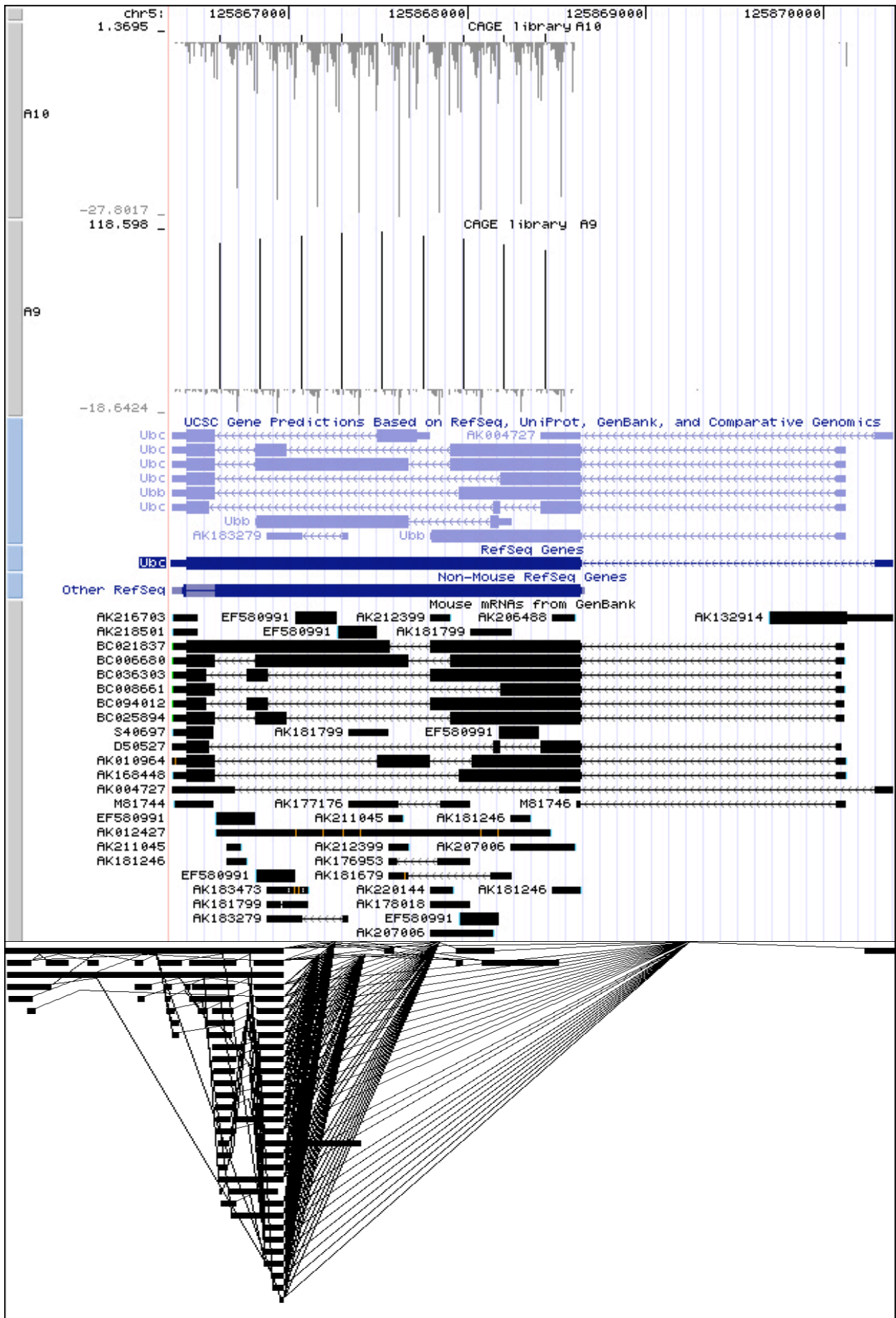
A further manual check of all the entries verified the reliability of the data and served to eliminate duplicated entries matching the same gene. A parallel further visual screening, which covered about 1-2% of the genes in the mouse genome, allowed to look for other possible interesting candidates acting in specific pathways.). Interstingly, one of the gene that has been found associated to an antisense overlapping transcript was Ubc, encoding for polyubiquitin chains, whose function may be also implicated in many aspects of the physiology and pathophysiology of the A9 –A10 neurons. Here is the scheme of the nanoCAGE tracks for Ubc.

Figure 3-47. nanoCAGE tracks for Ubc gene: in this case, vertical barswith negative tpm values represent sense transctpts, whereas positive tpm valuies are associated to antisense transcripts. UCSC gene predictions together with Swiss Bioinformatics alternative splicing, describe the presence of a very complex pattern of transcripts for this locus, which is confirmed by the multiple GenBAnk hits.

chr5:   125867000   125868000   125869000   125870000

1.3695 _

CAGE library A10

A10

-27.8017 _

118.598 _

CAGE library A9

A9

-18.6424 _

UCSC Gene Predictions Based on RefSeq, UniProt, GenBank, and Comparative Genomics

Ubc
Ubc   AK004727
Ubc
Ubc
Ubb
Ubc
Ubb
AK183279   Ubb

RefSeq Genes

Ubc

Non-Mouse RefSeq Genes

Other RefSeq

Mouse mRNAs from GenBank

AK216703	EF580991	AK212399	AK206488 AK132914
AK218501	EF580991	AK181799	
BC021837			
BC006680			
BC036303			
BC008661			
BC094012			
BC025894			
S40697	AK181799	EF580991	
D50527			
AK010964			
AK168448			
AK004727			
M81744	AK177176		M81746
EF580991	AK211045	AK181246	
AK012427			
AK211045	AK212399	AK207006	
AK181246	AK176953		
EF580991	AK181679	AK181246	
AK183473	AK220144	AK181246	
AK181799	AK178018		
AK183279	EF580991		
	AK207006		

221

Candidate non coding genes for establishing functional differences

From this zoom out view of the Ubc locus, the difference in the amount of sense and antisense transcripts can be better appreciated. They are the black bars on the line called (NC 000071-2396) and comprise at least two different transcripts, one overlapping within the 1st intron, and the other one overlapping at more 3`end region of the Ubc sense mRNA.
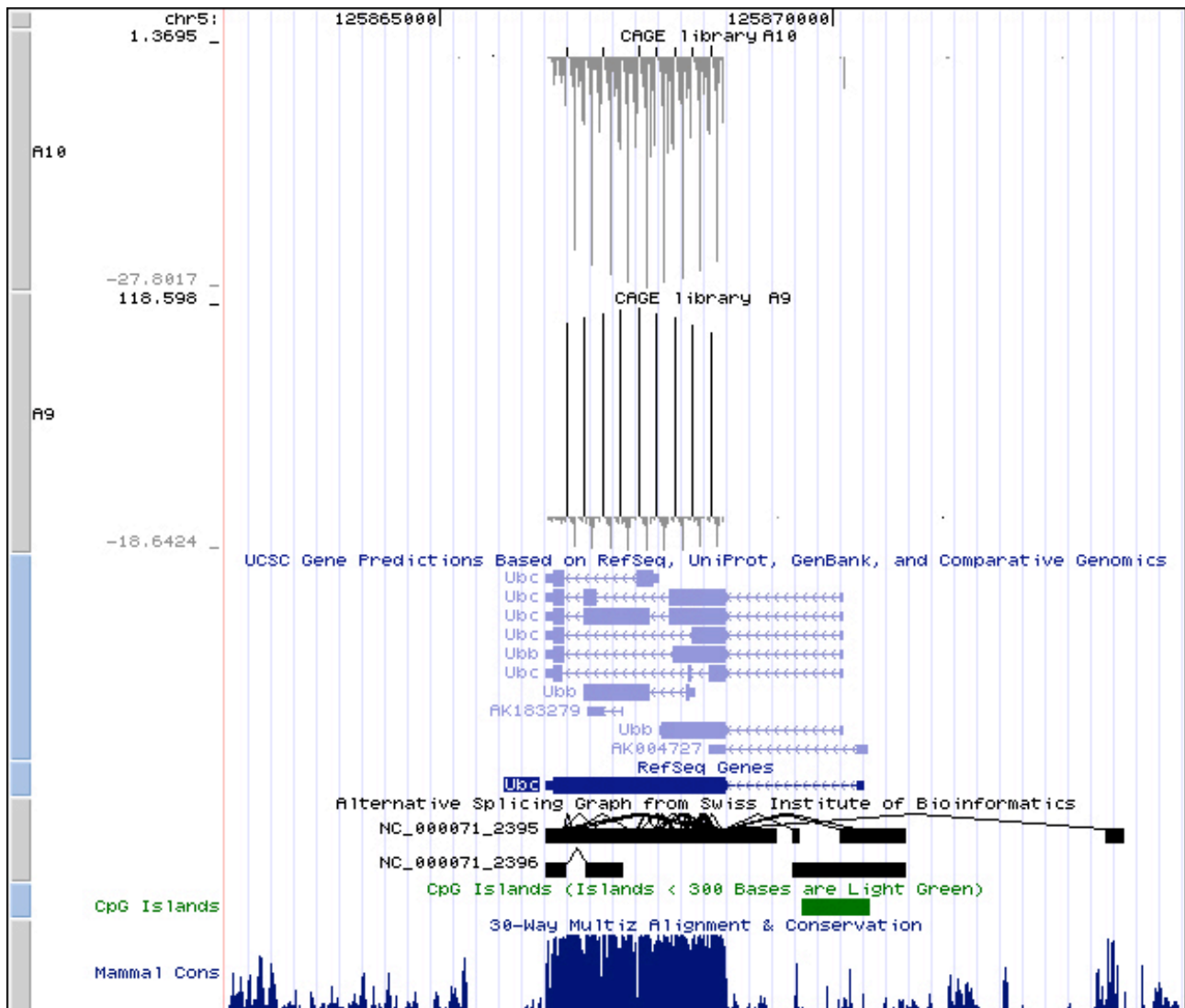

Figure 3-48

Interestingly, the predictions for antisense transcripts overlapping the sense Ubc mRNA, exhibit a complex alternative splicing pattern, which suggest an even further complex control in the transcriptional regulation of this locus.

The expression level of overlapping antisense RNAs at this locus is much higher in A9 compared to A10. To confirm that antisense transcripts do exist in vivo, an RT-PCR analysis has been done using specific combinations of oligonucleotides (figure 3-49).

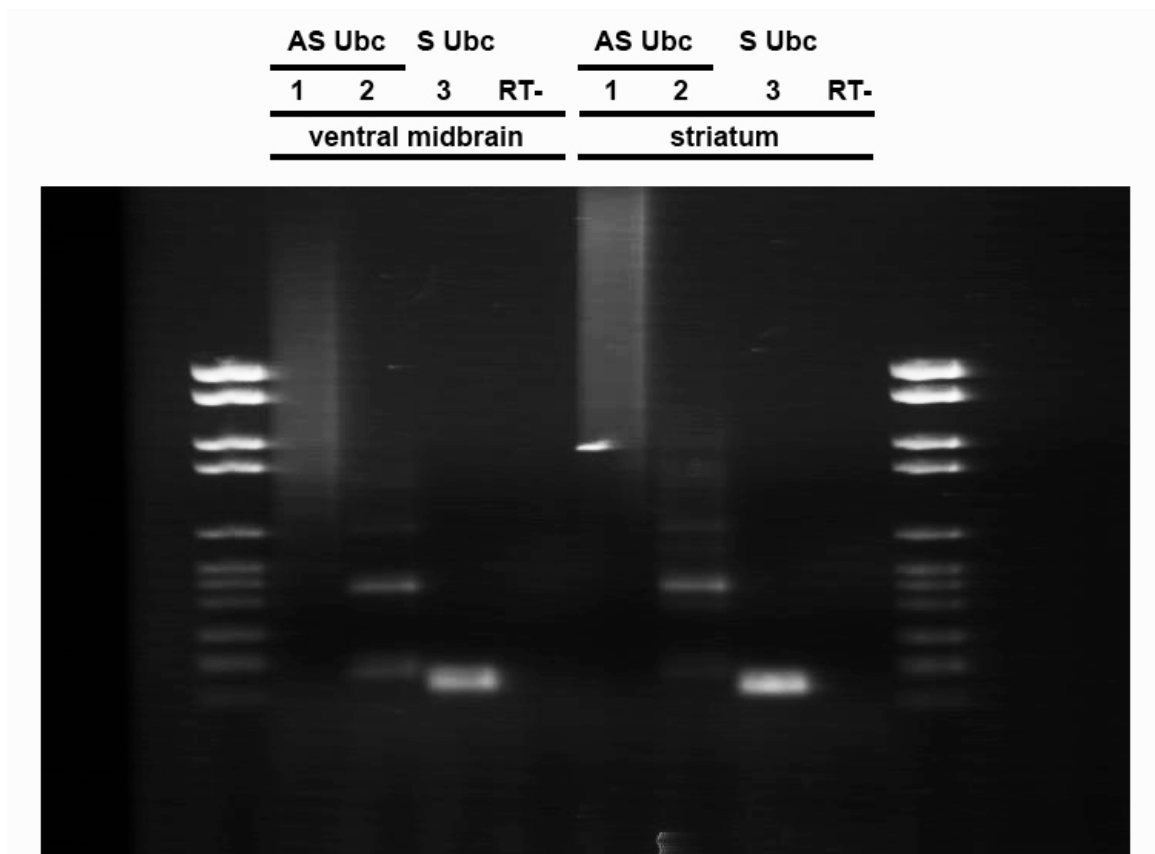In many independent biological and technical replicates, a smear of bands could be detected.

Figure 3-49. RT-PCR analysis of total RNA obtained from ventral midbrain and striatum respectively of the same C57Bl6 mouse. After a random primed RT, for the amplification of the specific cDNAs, the following comniations of primers have been used: (1) AS-Ubc rev/ AS-Ubc fwd; (2) AS-Ubc rev/ S-Ubc rev; (3) S-Ubc rev/ S-Ubc fwd. (see Oligonucleotide functional grouping for more details) The first combination of primers gave a much less clear smear , where a continuum of amplicons of different sizes are produced, whereas the combination (2) of primers gave a much more clear pattern of amplicons of the expected bands. Lane (3) represent a positive control for the expression Ubc of the sense transcript. In this case the primers S-Ubc fwd and rev were designed in a region at the 3`ends with no homology with the Ubb mouse gene, also coding for polyubiquitin chains.

Interestingly a more detailed analysis of the motifs present in the promoter of Ubc gene (considering 2 kb upstream and 5 kb downstream of the sense annotations) has been conducted using the Match informatics tool (BIOBASE), and found the presence of many motifs which seems to occur more than chance, for Pitx3 and Nurr1, which are both well known transcription factors controlling numerous genes involved in specifying and maintaining the DA-phenotype of the

mesencephalic DA-neurons. Here you can find all the genomic sequences screened with the motifs

is reported.

```
cttctacactaggtactttcttccttaccccagagccttgaggatctctt
cctttcaagctcaccttggccaggtccagggctcaagctttgcttcttcc
tcgcccctacccccaccccctacttctaaaatcaaggctgcctggctctga
atcttctagactctggaaaaaaacctgatgccacaaagctggtgactcag      S_AP-1
ccaagaaccaggaatcacgaggctgaagcaaagcgacagccaaatatagc     S_AGL-3
atggtgtggccggaagccagaatcactttagaaccaacttcagtgtttttc     S_Elk-1
ttcccttcctcaaacttctgtctgaattctagaaagaactgacctgccca
gaactctgcctcagaaatcactgtggcttcccagaaacccggagaatgac
attttaattttatagtggcatatgacctcgattcatggaaaaaaatcaca
gaaaattaagtctattctgaacaacacctgcagttaaagtatattttgaa
caatacctctgatcttgaatgtttatgaattcctaaaccagggagccaga     AS_Pax-4
caggcctgaattctcagcgtctgggtggtagaggcaggagtatcaactgc
aagttagaacccagcctgatctatgtcatgagtcccagggcagccagggc
tacatataggaaccctgtctcaaaaaaactcaaaaaatgctagattaaa     AS_Ptx-3
ggcatttaatagtgtgtgtccatctaagtgtgtgccacatgtatgtgaaa
gtgtccctagggacattggaccctcctgaagctggagtcacaagctagct
gcctgacaggttgctgggtattgaattcatgtactctggaaaaagacct
gttcctactggggagccatctctccccatcagggattcaattcttacctt
ctaaggtcttaggccttacatttcaaaaacctgagctgggcagtggtggc
gcaggcctttaatcccagcactttggaggcagaggcaggtggattctga    S_Bcd/Ptx-3
gttcgggggccagcctggtctacaaagtgagttccaggacagtcagggcta
cacagagaaaccctgtctcaaaaaaccaaaaaaaaaaaaaacaaaacaaaa
caaaacaaaaaaaaccctgaataatttgattcgtataagtgggatatatg
cttgctgtaatctgaacattgtactacacgtgcacatttgtgcttttttg     S_Ptx-3
tttttgtggggttttttgtttggttttttggtttttcaagtcaagtgtag     S_Nkx2-5
ccctggccattctggaacttgatctgtagaccatacttgggaggcagagg
caggaggacctatataaacttgagaccagcctggtctacatagcatattc
caggacaggcagggctacacagagaaacccccatctgaacaacaaaaatt
taaaaatcagggcttcagtgggttcagtctcagaatatgaacaacaaata
tgaacaacatttggaaattttgaccttttacttagtgacaagcacatgaa     S_Nurr1
aacatagaaccttctcgaatccaaggacacaaggtttaggaaaaatctgg     S_HSF
aaagatcaggttctgctgggcttcagatctaatgccaacagttttcaaga     S_CRP
actgtgtagactgtctgctttcccttgacaaactcagattcgttttttccaa   S_AG
atcggggagagagattccctccttcgtatttttttttttttaaatttattaaa
aacccttttgtcattttccttgttaggagatttgcacagtcatccaccctg
tcactgtgccaaaacagtgagggctgggcatagtgacccgccggggatga
ggggattgaggaggcgaggtcaccaatgggatcccagtgtgacaggagc
aaagccctgggtcccctagcgcggaaagaaaacagggggggtgggggt     S_Ptx-3
ggtggcctgactaggggcgccaggccgcgaagccgcagggcgcctgcgcc
aaggcccgctccggcctcagtgatcccagccgtgtttcgtgccgatcgt
CTCACGCGCGCTGATCCCTCCGCGGAGTCGCCCGAGGTCACAGCCCTGCC      <5`UTR>
CTCCCACACAAAGCCCCTCAATCTCTGGACGCCACCGTGAAACAACTCCG     S_Ptx-3
TGAGAGAGgtaccttgatagtttttagcctgtcgctttcgctgccgagact    <intron>
ggacccggcgttacaaagtagtccctgaccgcattgcccgcggaggacc
gcgcggaaggggggggggcggggcttcggtgactatataaagagacgccgg
gcgtgccgcagctagttccgtggagactgcgagttccgtctgctgtgtga
ggactgccgccaccaccgctggtgaggagaaagccgccgcacccggtcgg     S_Ptx-3
ggacgggaggctggaggcgagacggggcgagaggcagccccgcggcccag
acgtttggtttccgtggcccgcgcggaccgcggctgccccgaggcagagg
actgggcggcaagatggcggccagatggaagcctgaggggggaagacgcgg
ggctctgacgcgcaggacgaggttggggggaggaaaaaggcccgcgaggcc
gctgccctccggttaagccggggacgtcggagactgtggggtggggactg
aattaggggttgcgcgccgtaggagcctctgctgtgagagccgtggatatt    annotated AS
gggctggcccgagaggtcgattggcccggcgttcgtccgttcgtttgctg
aaagacggaagtgcgatcgagaccggaagggggttgggcggcggttcagc      S_NRF-2
ctgcctggcctgccgcccctgtacgtcgcgggttgcgtggcctcctaa
tggatagtgacgtcactatcttgactttagctttccctcggttgtaggac
agggtttgggtctcggcctccggtagcctctccagagtaaacaggaaccg
gaaattcagaggggaaatgtgagccattcttgtcctgtttcgttttaaga     S_Elk-1
atgtcgctgtacaactatgactactgaaacttttgggggggggttcgag
acgtttctctatgtagtcctggctgtcccgcatctcactcgtagatca
ggctggcctcgaacccagaaatcctcctgcctctgcctcccaagtgttgg
gacgaaaggcaccaccattgtcctgcgacaagggtgttttttttttaaact
gtcaaaatctctgcctctaccacccatgtgatgaggtccaaggccagta
ccaccactccagactaatttttaatcgttcagacaaaagtttggtgttctt
ttggggggaaggagagttgaggcaggattacactgtctctggctgaccttc     AS_Ptx-3
cagttagagatctgcccacctcagtgtccccagtgctgaggtcagcgata
ggcatgggctcagacttagttttgcagtagtaacttgctatattaccatt
ctgaaactgaatccgggactgctgtggtttcataacctcccagaggtcag     S_Pax-4
gctttctgcaaactgttcaaatagacagaaattgactttcagctgttgg
tatactgaagtctccatcctgtaaatttggtaatacaaaaagactcacca
tgccgaggtttcttaactttgttagtcaacaatcttattttcttgatggt
ttttcggggtggggggattggattcaagacagaatctgtgtagatagacc
```

```
ttgctatttagacttatagcatccagttgacaaatgttgatgccatccca
caaatatttgtgtgtcattcctgacctgtgaattgttttgtatattttgtga          ...S_BR-C Z1⇨
cagACGATGCAGATCTTTGTGAAAACCTTAACTGGTAAGACCATCACCCT
GGAGGTCGAGCCCAGTGACACCATTGAGAATGTCAAGGCAAAGATCCAGG
ACAAGGAGGGCATCCCCCCTGACCAGCAGAGGCTGATCTTTGCAGGCAAG          STRE↵
CAGCTGGAAGATGGCCGCACCCTGTCAGACTACAACATCCAGAAAGAGTC
CACCCTGCACCTGGTCCTTCGCCTCAGAGGTGGCATGCAGATCTTTGTGA↵
AGACCCTGACAGGCAAGACCATCACCCTGGAGGTCGAGCCCAGTGACACC          STRE↵
ATAGAGAATGTCAAGGCAAAGATCCAGGACAAGGAGGGCATCCCCCCTGA
CCAGCAGAGGCTGATCTTTGCAGGCAAGCAGCTGGAAGATGGCCGCACCC          STRE↵
TGTCAGACTACAACATCCAGAAAGAGTCCACCCTGCACCTGGTCCTTCGC↵          AS UBC forw
CTCAGAGGTGGGATGCAGATCTTTGTGAAGACCCTGACAGGCAAGACCAT
CACCCTGGAGGTCGAGCCCAGTGACACCATAGAGAATGTCAAGGCAAAGA
TCCAGGACAAGGAGGGCATCCCCCCTGACCAGCAGAGGCTGATCTTTGCA          STRE↵
GGCAAGCAGCTGGAAGATGGCCGCACCCTGTCAGACTACAACATCCAGAA
AGAGTCCACCCTGCACCTGGTCCTTCGCCTCAGAGGTGGCATGCAGATCT↵
TTGTGAAGACCCTGACAGGCAAGACCATCACCCTGGAGGTCGAGCCCAGT          AS UBC rev
GACACCATTGAGAATGTCAAGGCAAAGATCCAGGACAAGGAGGGCATCCC
CCCTGACCAGCAGAGGCTGATCTTTGCAGGCAAGCAGCTGGAAGATGGCC          STRE↵
GCACCCTGTCAGACTACAACATCCAGAAAGAGTCCACCCTGCACTTAGTC↵
CTTCGCCTCAGAGGTGGGATGCAGATCTTTGTGAAGACCCTGACAGGCAA
AACCATCACCCTGGAGGTCGAGCCCAGTGACACCATTGAGAATGTCAAGG
CAAAGATCCAGGACAAGGAGGGCATCCCCCCTGACCAGCAGAGGCTGATC          STRE↵
TTTGCAGGCAAGCAGCTGGAAGATGGCCGCACCCTGTCAGACTACAACAT
CCAGAAAGAGTCCACCCTGCACCTGGTCCTTCGCCTCAGAGGTGGGATGC↵
AGATCTTTGTGAAGACCCTGACAGGCAAGACCATCACCCTGGAGGTCGAG
CCCAGTGACACCATAGAGAATGTCAAGGCAAAGATCCAGGACAAGGAGGG
CATCCCCCCTGACCAGCAGAGGCTGATCTTTGCAGGCAAGCAGCTGGAAG          STRE↵
ATGGCCGCACCCTGTCAGACTACAACATCCAGAAAGAGTCCACCCTGCAC↵
CTGGTCCTTCGCCTCAGAGGTGGCATGCAGATCTTTGTGAAGACCCTGAC
AGGCAAGACCATCACCCTGGAGGTCGAGCCCAGTGACACCATAGAGAATG
TCAAGGCAAAGATCCAGGACAAGGAGGGCATCCCCCCTGACCAGCAGAGG          STRE↵
CTGATCTTTGCAGGCAAGCAGCTGGAAGATGGCCGCACCCTGTCAGACTA
CAACATCCAGAAAGAGTCCACCCTGCACCTGGTCCTTCGCCTCAGAGGTG
GGATGCAGATCTTTGTGAAGACCCTGACAGGCAAGACCATCACCCTGGAG
GTCGAGCCCAGTGACACCATAGAGAATGTCAAGGCAAAGATCCAGGACAA
GGAGGGCATCCCCCCTGACCAGCAGAGGCTGATCTTTGCAGGCAAGCAGC          STRE↵
TGGAAGATGGCCGCACCCTGTCAGACTACAACATCCAGAAAGAGTCCACC↵
CTGCACCTGGTCCTTCGCCTCAGAGGTGGCATGCAGATCTTTGTGAAGAC
CCTGACAGGCAAGACCATCACCCTGGAGGTCGAGCCCAGTGACACCATAG
AGAATGTCAAGGCAAAGATCCAGGACAAGGAGGGCATCCCCCCTGACCAG          STRE↵
CAGAGGCTGATCTTTGCAGGCAAGCAGCTGGAAGATGGCCGCACCCTGTC
AGACTACAACATCCAGAAAGAGTCCACCCTGCACCTGGTCCTTCGCCTCA↵
GAGGTCGCATGCAGATCTTTGTGAAGACCCTGACAGGCAAGACCATCACC
CTGGACGTCGAGCCCAGTGTTACCACCAAGAAGGTCAAACAGGAAGACAG          S UBC forw
ACGTACCTTCCTCACCACAGTATCTAAAAAGAGCCCTCCTTGTGCTTGTT
CTTGGGTGTGATGGGGGAGGTGTCTTAGTTTTCCCTATCTTTTAAGCTGT
TAACAAGTTTCATTGCACTTTGAATAAAGTTCTTGCATTCCAAAAAATCT          S UBC rev
TCAtttgtgttgttggttggttccttgtattgggatgtgtagagacttat
catatcaagtttgtgtcatcagtgtaacctatcaggtaagataatgttgg
actaggggactgtgtggcttttaaaagtattttccaaaaaccttgtttctc          AS_Ptx-3↵
agtgtgtagcctttgctgtcctggatcttcctctgtataccagtctgacc
tagaattcccaagaaatccacctgtcttttcaagtgctgggataaaaggta          ...S_Nkx2-5⇨
tttgccgggctggagatttggcgcaccagtgaagtgcactgacagctctt          ...S_E2F⇨
ccaggtcctgagttctagtcccagcaaccacatgttggttcataaccatc
tgtaataggatctgatgccctcatcggtgtatctgaagacagctagtgt
attaacatggaataaacaaatttaaaaagaaggtatttgccaccaccgcc          ...S_BR-C Z1⇨
ctgggtattgtaagattttaaagttattaggatgttggtattaatcta          S_Ptx-3⇨
tgaaggaggtcaacagaagcgttcagtcccttggtgtttgaaccaccatg
tggcctctgcctcccaagtgctggggattaaagtcgtgtgccaccatggcc          Bcd⇨/Ptx-3↵
ggcctaaatcttttttttttaaagattatttattttcacgtatgaatacact          ⇨
gttgctgtctttagacaacaccagaagagggcatcagatcccgttaaaga          ...S_Pax-6⇨
tggctgtgagccaccatgtggttgctgggaattgaactctgaacctctgg
aagagctaccagtgctcttaactactgagccatctcttcagccctgaact
taaatcttataagaggaacaaattctttgttctaaaatggcaatttaagg
tttaaacttaattttgcaagtgtatatttgcatgtatatgaatgctcaa          ...AS_Oct-1⇨
ggcagttgtcacattcccctagagataaggtcatttaagatcaatcaaac          S_Pbx-3⇨
tgggttctctggcaagttattatctgctttcactctgctgtgtctagccc          ...AS_HNF-1⇨
tgaagtggaattttctgtaaaacataccaggtaattatcacctagttgta
ctgaggacattatatagtaatggtgaaaagaagctgaagtgagcatggtg          S_Ptx-3⇨
gcactaactttttaatcctagcacctagcagaggcaagcagatctgagttc          S_Ptx-3⇨
atgagttccaggccagccaaggctgtagtaagacttgtcacacacaaaaa          ...S_Pax-4⇨
aatttccctaggttatgtgataacagctaaaggtcttggatggagtattg
acctgaatgttcccaagtgggtctagtcaggtgtgttaacgattggcctg          ...S_COMP1⇨
gagaagtgagtcagacattgtagtgctggtcttgtaattgacatactgtg
cctcagtgccacttgtccagagtggcccagtaagtaagtctagccttagc
atgtacctccactaaggattcaatgtatacgtagccatttcattaaaatt
```

```
gatactgaactggtcgtggtggcacatgcccttttaatcgccagcacttggga    S_Bcd/Ptx-3⟳
tgcagaggcaggtggatttattgaatttgagaccagcctggtctacagag
tgagttccaggacagccagagctacacaaagaaaccctgtctcgaaaaac
caaaaggaaaaaaaaaaaaaaaaactgataaaatgagtgtaatttgtcttg
gctatcctggaacttgtgccaccaaaccaggctggtttgattcatctatg
aggaaagtgcagggtattccatgaagacaatcctggactttgggaggaaa
tgcttgcttttttcttagttgctcatcctaagttagagctatgtaggcagc
aatatgatttgatgaccctgtttctttgtacctgttctaagttgttggtt
ttgttttgttttttttaaagatggtgaggaataatgaaataaccagcctt
agccatgtaagcacatgatctgctctggagctaaaattcacctcactttg
ttttttttaattctagttctgttcagggagaacacatggtgtagtccagac
tagctttgaacttgtatgttcctgcctcaacttccctaatgtcagattta
taagctcggaggttaagcctgagcttaagggtttctccaacagaccaagg
ttcaatttcttgtacccatatcacagatcgtttgtaatgccatacctct    S_Ptx-3⟳
ttggacttttgggcagaaggcaggcaaaacaatacaaagattttaaagtt
aaaagtatagcctgttggtgagtgtgttccagaggcaggcaaagttaagg
ccagcctggtattcaagactgaatttcaggatatccagggctacacagag
aaaccattaaaattagaagcatggttgccatgctcaacacaagcaagaat    AS_Ptx-3↵
tttttttttttacatttatttgtggctaagcaattgctgttctctagccc
ttagcccaggatcccaccacagttgaaattaaaccacaacgcctactggg
ttacagcaggtaagaggtcccaaccattccccaagtctgtctacatgtgt
gaagattagagtcagttttttccactgtgtggttggtacccaaagctccat    AS_Ptx-3↵
atcggtctagcagcaagcacttgctgacctgttgctcatagtgtctgaat
gtgtgctacttggtttcaaatactcattttataatggaacaatgccatc    S_Ptx-3⟳
cttgcttttggctggtgctcctaccccattgtgttcagcttcatacatac
aggtcctcagagacactgggagagctctaggctatacactgactactgca
tgttttcacttcagatctcggtattattaggtctattagccttacaacat
atacaaaggggagaaattccttttctcaaaaacagaaagccagggtaaaa    S_Ptx-3⟳
gaactggactgtcaggagcatatgaagctcggcttcctagagtcctcttt
ctagagtcctcggtataatgaacagagtttaaagtactccatagtttcac
ttcacacttggaactaagttcacttacagaggtgaggttgagtcagaaag
tactaggcctgagtgtatatactaaccacttgtaactcaagctttaaggg
gagtcaacagcaggcacctgcactcaaatgcccatacttccagatgggca
gtcctggcaagcaatactctcccagaagaggacatcactgggccacatgc
acgtttagtaatgaaggctgcagtgcccccaagaggttgtagatggttaa
gcaattgctgttcttagccaagggtcccaccacagttgaaattaaaccac
aacacccactgaattataccaagcaaaaggtctcaaccactccacagatt
acccacattctgtaaagcctcccccacaatttttcctttacttcacccaagt    S_Ptx-3⟳
tgtccttgttcctaaggccactgcacctacttaggtgtttcctctcctaa
ttttttttccccacagagtgtaccttgtaacccaggctgaccttgaacctc
cttatgcctcagtgttgtgtaggactgttttacaagccacactgccacca
tcttcatcaaatcagctgggccagcttgcaaaagtaacagggcctgttga
ctattggtattgtttttgtaggagaggcaggcagggtcatgggaccctcaa
ctgaatatccagccagttctcccaactttctcataactgtatatagtctc
caggaggggctagagaaaataggcaaaaaagaactaggagaagcaggctt
tatgtgaccacagggaagacgtcatccatgctgggtacagactttacagt
gtcattgctttagcggtcacccacgcttctgctcataaccctcacatat
gatctgtaagtaagcccagtaaacttactcgtttcagtgagcaaactttg
gtggactcaagccaacatttgtccttggaccttaggaggaggggggagatg
tggtttccgtttccccaagggaagggatttttagcaacattgagtttccag
gtgttaagagtcctttctttattctagtggcttaacattgggtagtcacc
catggtgacaatgacacttgtagctctaggtggttcaggtgctaaggtgaa
gtgctgcaaaactaaagagcctcagaggatggagccagggaccaggtgct
gggagagaacttgggtgtagatcaacaggagacctcactttcaaaagggg
ccaaatgtttggtgtggagcagggcctagacactggtttttctgtagtgct
tggaatgggacccaggacccttaaacagctgagagtttgaaaatctaggga
attgattgattgactgattgatcctgcagctggcactggaaactctgagt
gaacaccgggtacactgaagaatgatgatctcagactaaggcccagagtg
tagagcactaactagtagtaaagccatgaggcacagggttctatctcttg    S_Ptx-3⟳
cacagcataagtgggcccccgtggtgcaggcctgcagttccagagagggag
ctgaaggctggaggtttgaggacaactggaccacactagaccctatatca
tcactaatgacttattgagactgtcttttgtccaaagtatgtgaaattaa
cattttctcccagtagggttcattttgtggttgttcaaccaaagctttgt
tgagctgctactcagaggctgctccagtagaggaaaatcaaaacatagag
acccaaatctgctatgtgatggtgtttgaagctcctgtctctgttctcat
caatacagctggtctgttggagatgtgccttgcaaacatcaatgcccaga
aaaattaaggtgttacctctttcctgtgacttcctggaggtggaaactca
agctgacttttttccctcattactcttctttctcctctgtgttaaggacat
gaattttggtaggaggtgaagattagggtatacctgtctccccttagaga    AS_Ptx-3↵
tcatagaaaaccatggctcaggccaggggagggggtacactaagctcagca
ctgaagaggtagaaaggagggccatcagtttagaccatctgggctacaaa
atg   (5kb downstream)        RED AS tags & RT primers
                              GREEN S tags & RT primers
```
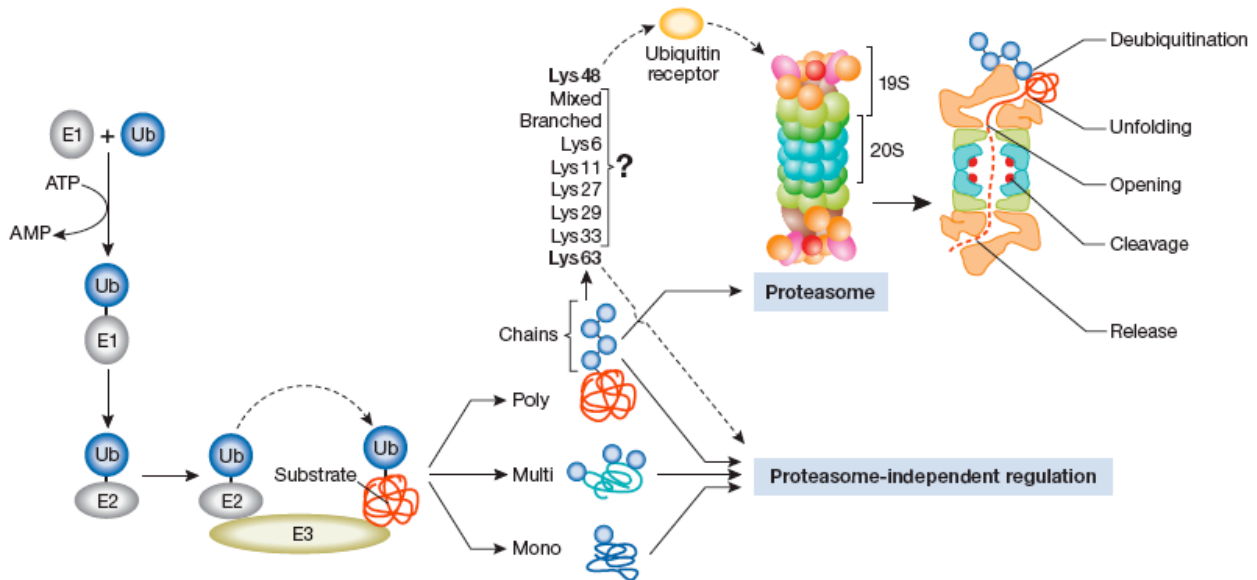
Fig 3-50. Ubiquitination. Ubiquitin (Ub) is activated by E1, transferred to E2s, which bind to E3 for conjugation of Ub predominantly onto lysine residues in substrates; E3s confer substrate specificity to the system. Additional rounds of Ub conjugation onto substrate-attached Ubs lead to the formation of Ub chains. Different lysines in Ub can be used for chain formation. Chains linked through lysine in position 48 (Lys 48) in Ub are the predominant proteasome-targeting signals *in vivo*. Lys 63-linked chains function in a proteasome-independent manner. The function of other chain topologies *in vivo* is unclear. Figure taken from Kaiser and Fon, 2007.
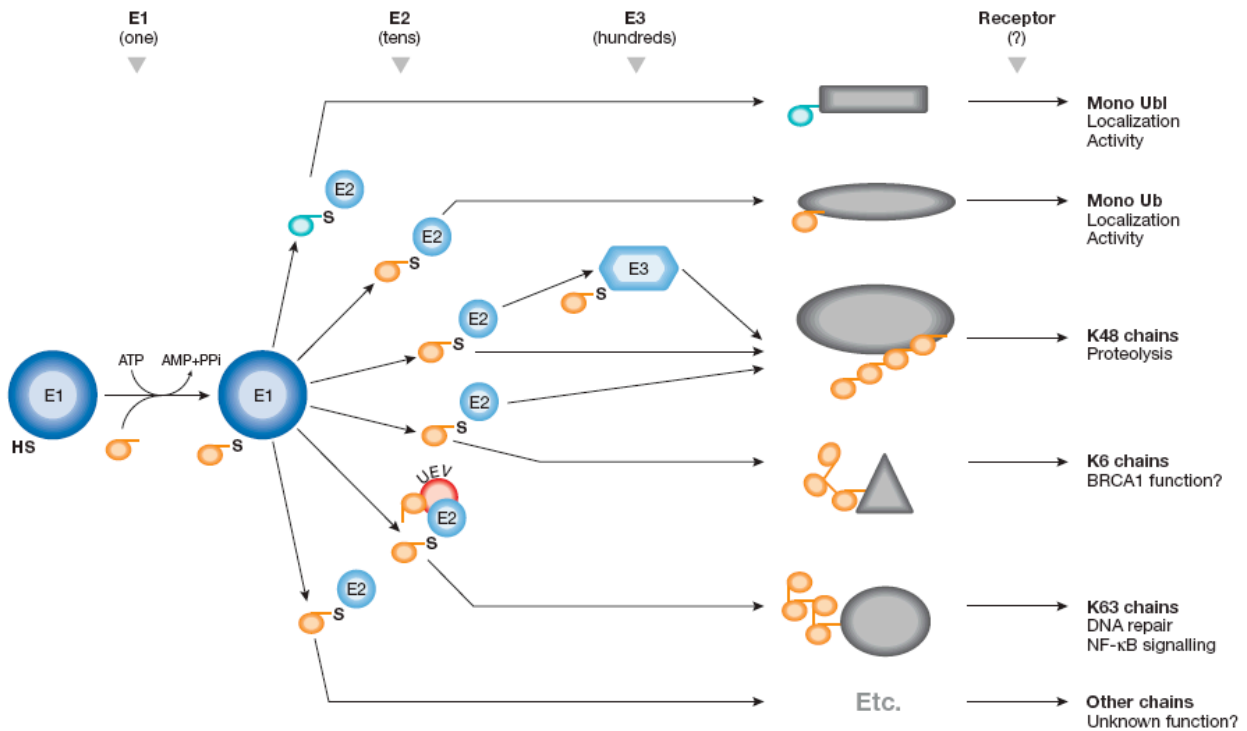

Figure 3-51. | Assembly and fates of various forms of the ubiquitin domain. The E1 activating enzyme adenylates the carboxyl terminus of ubiquitin (Ub) and then forms a thiolester with the E2 conjugating enzymes that act as mobile carriers of activated Ub. Ub ligases are responsible for the specificity of attachment of Ub to the target protein through the recruitment of both an E2 thiolester and a specific substrate. Modification by a single Ub domain regulates localization and/or activity of conjugated proteins. PolyUb chains can be formed with different linkages and these direct proteins to different fates, presumably requiring chain-specific receptors. Note that more than one E2 can work with a given E3 and that several E3s can use a single E2. BRCA1, breast cancer 1; UEV, ubiquitin-conjugating enzyme E2 variant protein. Picture taken from Wilkinson et al 2005.

SECTION 5. DISCUSSION and CONCLUSION

Among all the biological systems, neural networks in the CNS probably has one of the most complex organization in terms of their diversity in properties and function as well as their interconnections. There is an underestimated hetereogeneity among neurons.

Only a few morphologically defined cell types express usefully distinctive cytochemical markers, and those sharing common markers do not necessarily display the same patterns of connections or electrophysiological properties.

To overcome the limits inherent to a classical anatomical description, in the last years, multiple methods and approaches have been used to gain insights into the molecular complexity which characterizes neuronal populations. The use of high-throughput techniques for genome-wide expression profiling such as microarray, provided invaluable tools for a deep analysis of the transcriptomes of specific subset of neurons, opening new possibilities in the field of functional genomics. However, for long time, the need for a sufficient amount of RNA to be used as template for subsequent analyses, strongly limited the application of these techniques, preventing their use for particularly small biological samples such as neuron subpopulations in the brain. To overcome these limitations, we first applied a two rounds amplification using method SMART7 to obtain a sufficient amount of cDNA to be used as fluorescent-labeled probe in a classical cDNA microarray experiment, which normally requires micrograms of starting material. The SMART7 protocol was applied to profile decreasing amounts of RNA obtained from serial dilutions of culture striatal neurons, and it has been shown to produce reliable results even in the range 1-10 ng.

However results obtained from microarray analysis are biased towards the set of used probes, depending on each particular platform used in the experiment. To bypass these inherent limitations of the microarray, we took advantage of new high-throughput sequencing technologies, to adapt a classical tag-based method for genome-wide TSS discovery to be applied to small group of cells.

In a collaborative effort with dr. Piero Carninci from Genome Science Center at RIKEN (Japan), we developed nanoCAGE, based on Cap Analysis of Gene Expression (CAGE), as a simple method to analyze the gene network in small samples containing 10-50 nanograms of total RNA. Few thousands DA cells were isolated using laser capture microdissection (LCM) technology by Dr. Christina Vlachouli from two different subpopulations (A9 and A10) of mouse midbrain from a transgenic TH-GFP mouse, in which a reporter gene is expressed in all catecholaminergic neurons. The RNA obtained from these highly selected neuronal populations was then used as starting material for a nanoCAGE profiling, which allowed an unbiased analysis of the promoters activity along the entire genome. After mapping and clustering the tag sequences on the mouse genomic DNA, a statistical analysis has been employed to uncover new interesting genes for subsequent experimental validation.

The new nanoCAGE method has been proved to give informations about both the quantitative profile of gene expression and the TSS usage frequency. This allowed a complete description of the promoters used by a particular cell type, taking advantage of the depth of coverage reachable using new high-throughput sequencing platforms such as Illumina-Solexa.

It has been calculated in fact that only having millions of tag sequences for experiment can ensure an almost complete coverage of the all RNA transcripts present in a cell (Zhu et al 2008).

The usage of random primer in the reverse transcription step for cDNA synthesis gave an invaluable information about expressed non coding RNAs, which can often be devoid of a poly(A) tail and it is particularly useful for detecting antisense transcripts, often missed by the microarray analysis. The major advantage of nanoCAGE compared to all other methods was the detection of differential cell-type specific promoter usage.

We were able to detect new unannotated transcript isoforms, even in the case of widely studied genes, such as α-synulein, dopamine transporter, vescicular monoamine transporter and catechol-O-methyltransferase, all of which have been subject to extensive research investigations, due to their central role in the physiology and pathophysiology of dopaminergic cells.

Most of the research conducted on these genes has considered the RefSeq annotations for their transcripts and corresponding proteins as reference ascertained points on which all the experimental framework has been based. But there are accumulating evidences proving that the current transcript annotations have to be revised and reconsidered in the view of more recent genome-wide investigations, which gave convergent results even using independent different platforms and systems such as RACE, genome tiling array, CAGE and more recently RNA-Seq (Carninci et al 2005, Gustincich et al 2006, Sandelin et al 2007, Denoeud et al 2007, Cloonan et al 2008, Sultan et al 2008). These new unbiased analyses revealed that most of the genes in the genome have more than one associated transcript. Many genes are transcribed from new distal or internal TSSes, not comprised in the current annotations, which can either give rise to regulatory non coding RNAs or to potentially coding transcripts often producing shorter proteins (Denoeud et al 2007, Oyama et al 2007). This hidden layer of transcripts contributes to the complexity of gene network regulatory dynamics in a cell (Yasuda and Hayashizaki 2008).

Studying the midbrain dopaminergic cell system using an unbiased high-throughput approach together with single cell harvesting by LCM, we uncovered at least part of this hidden complexity to unveil that some of the genes involved in the dopamine homeostasis are actually transcribed from alternative promoters.

The observation that cell-type specific alternative promoters are differentially used between A9 and A10 neurons in normal physiological conditions. may be important on our knowledge of the biology of these neurons. This may open new working hypotheses to explain, at least in part, the differential vulnerability of these two cell populations to environmental insults and their differential responses to pathological conditions. As previously described in section 4, all of these genes, α-synuclein, Dat, Vmat2 and Comt, are involved at different levels, in establishing and regulating dopamine homeostasis inside the cell. Whereas for at least three of these genes (Dat, Vmat2 and Comt), it is clear and well established a role in dopamine metabolism, uptake, storage and

degradation, for α-synuclein, a normal physiological function inside the cell is not as much clear as one would expect from the large number of studies focused on this protein.

Since the first times when Parkinson`s disease was recognized as a neurodegenerative condition mainly affecting the dopaminergic neurons, it became clear that there is a difference in the susceptibility among distinct neuron subpopulations: A9 neurons in the ventral tier of SN are the most affected group, whereas the A10 VTA neurons are relatively spared.

However, the precise reasons for this different vulnerability are not known. It has been largely accepted that Lewy bodies, mainly composed of α-synuclein aggregates and other polyubiquitinated proteins, are almost invariantly found to be associated to PD surviving neruromelanin-positive SN neurons.

How α-synuclein may be implicated in the pathophysiological process of PD is far from obvious. A puzzling aspect of α-synuclein -mediated cytotoxicity and LBs formation with regard to PD is the preferential and selective neurodegeneration of dopamine-producing neurons, such as the substantia nigra. α-Synuclein is ubiquitously expressed at high levels in virtually all regions of the brain, where it is believed to account for up to 0.1% of total brain proteins (Clayton et al 1998). If high expression levels of α-synuclein were the only cause of protofibril formation and LBs, then most brain regions could be expected to have aggregates of α-synuclein.

Yet in PD the SN neurons, which express considerably less α-synuclein than other areas of the brain, are specifically affected, whereas other brain regions expressing higher levels of α-synuclein are relatively spared. Increasingly then, the evidence suggests the likelihood that the specific vulnerability of these neurons is linked to other factors, the prime candidate being dopamine itself, which necessarily can confer the selectivity for these neurons.

Multiple evidences, both direct and indirect, have been recently accumulating showing a close relationship among α-synuclein and dopamine, whereby this protein probably participates in regulating the biosynthesis, storage into vescicles, release as well as reuptake of the

neurotransmitter, thus implicating α-synuclein as a central player and regulator of the dopamine homeostasis in those neurons.

Dopamine is inherently unstable and can generate reactive oxygen species (ROS) and, via monoamine oxidase (MAO), H_2O_2 (Lotharius et al 2002). Dopamine is synthesized in the cytosol and rapidly pumped by vesicular monoamine transporter 2 (VMAT2) into synaptic vesicles, where the low vesicular pH and the absence of MAO limit its breakdown. Defects in the early secretory pathway could cause a shortage of synaptic vesicles and reduce delivery of VMAT2 to the synapse. This would impede dopamine loading and increase cytosolic dopamine levels and consequently ROS production. Although dopamine itself may not be toxic, dopamine metabolites play a role in α-syn aggregation (Xu et al 2002, Galvin 2006) and dopamine- α-syn adducts stabilize α-synuclein protofibrils (Conway et al 2001). It has been shown, that, dopamine can have an opposite effect on α-synuclein conformation depending from its redox state: dopamine istself normally is tought to induce the aggregation of α-synuclein in large fibrillar structures (Moussa et al 2008), whereas the oxidized metabolites of dopamine inhibit the fibrillization process stabilizing the protofibrills or toxic olgomeric species (Conway et al 2001). Therefore, dopamine can directly influence the conformation, hence the equilibrium among the membrane-bound form and the soluble cytoplasmic form of α-synuclein protein. More recently many studies have linked α-syn to the function of other key genes implicated in dopamine homeostasis: α-syn normally modulates the enzymatic activity of TH preventing its interaction with kinases (Perez et al 2002). On the contrary other chaperone proteins 14-3-3, favor TH activity by enhancing TH phosphorylation by calmodulin dependent kinases and ERKs (Ichimura et al 1988, Toska et al 2002). However it has been shown that aggregated α-syn while decreasing the total content of TH also stimulates its serine phosphorylation, suggesting that excessive aggregated α-syn is no longer able to inhibit TH (Alerte et al 2008).

Morover α-syn has been shown to inhibit the activity of ERK through its direct binding to the MAP kinase (Iwata et al 2001a, Iwata et al 2001b), an effect also seen with the A53T mutant (Iwata et al

2001b). This inhibitory effect of α-syn can be extended to other kinases such as calcium-calmodulin-dependent kinases (Lee and Lee 2002) and PKC (Okochi et al 2000) and contributes to counteract the TH activity.

The NAC region of α-syn has been also shown to directly interact with the C-terminal portion of DAT (Wersinger et al 2003), suggesting for the first time that α-syn can also play a direct role in the regulation of dopamine uptake process.

DAT is a major determinant of dopamine homeostasis and synaptic strength since it is the primary mechanism by which endogenous neurotransmitter is rapidly removed from the synaptic cleft. Increases or decreases in DAT function will concomitantly decrease or increase synaptic dopamine concentrations, respectively, thereby regulating the activity of multiple post- and presynaptic dopamine D1- and D2-like receptors. The DAT is also the sole means by which the parkinsonism inducing neurotoxin MPP^+ is transported into nigral neurons (Gainetdinov et al 1997, Takahashi et al 1997). The importance of DAT function in the control of synaptic availability of DA suggests that its own regulation may be a crucial component in the maintenance of dopaminergic neurotransmission, since enhanced DAT activity would not only decrease extracellular levels of DA but also increase intracellular levels of DA upon reuptake.

The principal means by which DAT function is regulated is through the rapid shuttling of DAT to and from the plasma membrane. Previous work in cellular expression systems has shown that reduced cellular dopamine uptake mediated by DAT may be directly correlated to the possible phosphorylation of DAT by kinases, causing rapid redistribution and internalization of DAT away from the plasma membrane (Pristupa et al 1998, Melikian and Buckley 1999, Daniels and Amara 1999). The participation of kinases such as PKC that are dynamically activated by changes in intracellular $[Ca^{2+}]$ levels, suggests rapid adaptations of the neuron in response to signal transduction-induced changes in calcium flux to attenuate DAT function. Recent data have shown that MAP kinases ERK1 and 2 phosphorylate DAT upon signal transduction, increasing the amount of DAT at the plasma membrane and dopamine reuptake (Moron et al 2003). Since α-syn binds to

MAPK and tends to inhibit its activity, as shown for the regulation of TH enzyme, it is likely that the presence of α-syn at the nerve terminals tends to blur MAPK activity and decrease the amount of DAT inserted in the membrane of nerve terminals. Sidhu and colleagues have shown that upon coexpression in fibroblasts or in basal conditions in mesencephalic neurons, α-syn tends to markedly decrease by 30– 50% the activity of DAT by reducing the reuptake of dopamine (Wersinger and Sidhu 2003, Wersinger et al 2003). The reduction in DAT activity was due to a decrease in DA uptake velocity by the transporter without any change in DAT expression levels.

A consequence of the α-syn -mediated attenuation of DAT activity was that upon exposure of cotransfected cells to DA, there was a diminished DA-induced oxidative stress and cytotoxicity, suggesting that disruption of the ability of α-syn to regulate transporter function may be one of the most important determinants in the genesis of dopaminergic neurodegeneration.

The cause that initiates the dysregulation of α-syn-mediated DA homeostasis is currently not known. It has been observed that a transcriptional dysregulation of α-syn is a common event since gene dosage seems to be crucial for regulating its activity. α-syn gene duplications and triplications lead to disease, and the same effect is observed when α-syn is overexpressed in transgenic animal models od PD.

Interestingly, we have preliminary evidences that different subpopulations of mesencephalic DA cells amy use alternative promoters for α-syn and DAT. Currently, nobody knows what is the relative contribution that alternative promoters and different isoforms of the protein may play in the regulation of normal DA homeostasis and in the pathogenesis.

With the development of a new high-throughput genome-wide technique (nanoCAGE) to study the promoters activity in few isolated homogeneous groups of neurons, we were able to demonstrate that a fine-tuning in the transcriptional regulation do exist among A9 and A10 neurons *in vivo*, and in particular the usage of cell-type specific alternative promoters might account for slight differences in the internal homeostasis control in those neurons, which may give some hints about the selective A9 vulnerability, and the relative resistance to stress of VTA A10 neurons.

The fact that a cell-type specific transcriptional control seems to regulate in different ways the DA-phenotype associated to these distinct neuronal populations, could change our current perspective, further confirming that those two neuronal populations have indeed much more differences than previously thought. These observations could also open new questions about the physiological function associated to each specific isoform of those genes *in vivo*.

Since all of the genes on which we focused our attention are actually pharmacological targets in the cure of PD, we plan testing how a selective modulation of the transcription of each isoform of those genes could affect the normal DA-homeostasis and/or the their dysfunction which leads to develop a neurodegenerative condition.

For instance, an isoform specific targeting with antisense oligonucleotides or transcriptional interference (Janowski et al 2006) would be one possibility to explore in the next future.

Acknowledgements

References

Abdolmaleky, H.M., Smith, C.L., Zhou, J.R., and Thiagalingam, S. (2008). Epigenetic alterations of the dopaminergic system in major psychiatric disorders. Methods Mol Biol *448*, 187-212.

Abdolmaleky HM, Thiagalingam S, Wilcox M. (2005). Genetics and epigenetics in major psychiatric disorders: dilemmas, achievements, applications, and future scope. Am J Pharmacogenomics. *5*, 149-60.

Adams, K.A., Maida, J.M., Golden, J.A., and Riddle, R.D. (2000). The transcription factor Lmx1b maintains Wnt1 expression within the isthmic organizer. Development *127*, 1857-867.

Agid, Y. (1991). Parkinson's disease: pathophysiology. Lancet *337*, 1321-24.

Akam, M.E., Martinez-Arias, A., Weinzierl, R., and Wilde, C.D. (1985). Function and expression of ultrabithorax in the Drosophila embryo. Cold Spring Harb Symp Quant Biol *50*, 195-200.

Al-Shahrour, F., Minguez, P., Tárraga, J., Medina, I., Alloza, E., Montaner, D., and Dopazo, J. (2007). FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. Nucleic Acids Res *35*, W91-96.

Alavian, K.N., Scholz, C., and Simon, H.H. (2008). Transcriptional regulation of mesencephalic dopaminergic neurons: the full circle of life and death. Mov Disord *23*, 319-328.

Albin, R.L., Young, A.B., and Penney, J.B. (1989). The functional anatomy of basal ganglia disorders. Trends Neurosci *12*, 366-375.

Alerte, T.N., Akinfolarin, A.A., Friedrich, E.E., Mader, S.A., Hong, C.S., and Perez, R.G. (2008). Alpha-synuclein aggregation alters tyrosine hydroxylase phosphorylation and immunoreactivity: lessons from viral transduction of knockout mice. Neurosci Lett *435*, 24-29.

Allen, N.C., Bagade, S., McQueen, M.B., Ioannidis, J.P., Kavvoura, F.K., Khoury, M.J., Tanzi, R.E., and Bertram, L. (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. Nat Genet *40*, 827-834.

Altman, J., and Bayer, S.A. (1981). Development of the brain stem in the rat. V. Thymidine-radiographic study of the time of origin of neurons in the midbrain tegmentum. J Comp Neurol *198*, 677-716.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389-402.

Alves Da Costa, C., Paitel, E., Vincent, B., and Checler, F. (2002). Alpha-synuclein lowers p53-dependent apoptotic response of neuronal cells. Abolishment by 6-hydroxydopamine and implication for Parkinson's disease. J Biol Chem *277*, 50980-84.

Ambros V., Lee R.C., Lavanway A., Williams P.T., Jewell D. (2003). MicroRNAs and other tiny endogeneous RNAs in C. elegans. Curr Biol. *13*, 807-18.

Andén, N.E., Hfuxe, K., Hamberger, B., and Hökfelt, T. (1966). A quantitative study on the nigro-neostriatal dopamine neuron system in the rat. Acta Physiol Scand *67*, 306-312.

Andersen A.A., Panning B. (2003). Epigenetic gene regulation by noncoding RNAs. Curr Opin Cell Biol. *15*, 281-9.

Angrist, B., Peselow, E., Rubinstein, M., Wolkin, A., and Rotrosen, J. (1985). Amphetamine response and relapse risk after depot neuroleptic discontinuation. Psychopharmacology (Berl) *85*, 277-283.

Anney RJ, Rees MI, Bryan E, Spurlock G, Williams N, Norton N, Williams H, Cardno A, Zammit S, Jones S, Jones G, Hoogendoorn B, Smith K, Hamshere ML, Coleman S, Guy C, O'Donovan MC, Owen MJ, Buckland PR. (2002). Characterisation, mutation detection, and association analysis of alternative promoters and 5' UTRs of the human dopamine D3 receptor gene in schizophrenia. Mol Psychiatry *7*, 493-502.

Arce L, Yokoyama NN, Waterman ML. (2006). Diversity of LEF/TCF action in development and disease. Oncogene *25*, 7492-504.

Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P., and Proudfoot, N.J. (1997). Intergenic transcription and transinduction of the human beta-globin locus. Genes Dev *11*, 2494-2509.

Asthana, S., Noble, W.S., Kryukov, G., Grant, C.E., Sunyaev, S., and Stamatoyannopoulos, J.A. (2007). Widely distributed noncoding purifying selection in the human genome. Proc Natl Acad Sci U S A *104*, 12410-15.

Avantaggiato, V., Acampora, D., Tuorto, F., and Simeone, A. (1996). Retinoic acid induces stage-specific repatterning of the rostral central nervous system. Dev Biol *175*, 347-357.

AXELROD, J., and TOMCHICK, R. (1958). Enzymatic O-methylation of epinephrine and other catechols. J Biol Chem *233*, 702-05.

Ayoubi, T.A., and Van De Ven, W.J. (1996). Regulation of gene expression by alternative promoters. FASEB J *10*, 453-460.

Babak, T., Blencowe, B.J., and Hughes, T.R. (2005). A systematic search for new mammalian noncoding RNAs indicates little conserved intergenic transcription. BMC Genomics *6*, 104.

Baek D, Davis C, Ewing B, Gordon D, Green P. (2007). Characterization and predictive discovery of evolutionarily conserved mammalianalternative promoters. Genome Res *17*, 145-55.

Bai L, Santangelo TJ, Wang MD. (2006). Single-molecule analysis of RNA polymerase transcription. Annu Rev Biophys Biomol Struct. *35*, 343-60.

Bains W, Smith GC. (1988). A novel method for nucleic acid sequence determination. J Theor Biol. *135*, 303-7.

Baker, H., Kobayashi, K., Okano, H., and Saino-Saito, S. (2003). Cortical and striatal expression of tyrosine hydroxylase mRNA in neonatal and adult mice. Cell Mol Neurobiol *23*, 507-518.

Ball, P., Knuppen, R., Haupt, M., and Breuer, H. (1972). Kinetic properties of a soluble catechol O-methyltransferase of human liver. Eur J Biochem *26*, 560-69.

Ballmaier, M., Zoli, M., Leo, G., Agnati, L.F., and Spano, P. (2002). Preferential alterations in the mesolimbic dopamine pathway of heterozygous reeler mice: an emerging animal- based model of schizophrenia. Eur J Neurosci *15*, 1197-1205.

Baptista, M.J., O'Farrell, C., Daya, S., Ahmad, R., Miller, D.W., Hardy, J., Farrer, M.J., and Cookson, M.R. (2003). Co-ordinate transcriptional regulation of dopamine synthesis genes by alpha-synuclein in human neuroblastoma cell lines. J Neurochem *85*, 957-968.

Barbour, V.M., Tufarelli, C., Sharpe, J.A., Smith, Z.E., Ayyub, H., Heinlein, C.A., Sloane- Stanley, J., Indrak, K., Wood, W.G., and Higgs, D.R. (2000). alpha-thalassemia resulting from a negative chromosomal position effect. Blood *96*, 800-07.

Barker, R.A., Dunnett, S.B., Faissner, A., and Fawcett, J.W. (1996). The time course of loss of dopaminergic neurons and the gliotic reaction surrounding grafts of embryonic mesencephalon to the striatum. Exp Neurol *141*, 79-93.

Baugh, L.R., Hill, A.A., Brown, E.L., and Hunter, C.P. (2001). Quantitative analysis of mRNA amplification by in vitro transcription. Nucleic Acids Res *29*, E29.

Bauman, A.L., Apparsundaram, S., Ramamoorthy, S., Wadzinski, B.E., Vaughan, R.A., and Blakely, R.D. (2000). Cocaine and antidepressant-sensitive biogenic amine transporters exist in regulated complexes with protein phosphatase 2A. J Neurosci *20*, 7571-78.

Beach, T.G., Sue, L.I., Walker, D.G., Lue, L.F., Connor, D.J., Caviness, J.N., Sabbagh, M.N., and Adler, C.H. (2007). Marked microglial reaction in normal aging human substantia nigra: correlation with extraneuronal neuromelanin pigment deposits. Acta Neuropathol *114*, 419-424.

Beatty, L., Weksberg, R., and Sadowski, P.D. (2006). Detailed analysis of the methylation patterns of the KvDMR1 imprinting control region of human chromosome 11. Genomics *87*, 46-56.

Beckstead, M.J., Grandy, D.K., Wickman, K., and Williams, J.T. (2004). Vesicular dopamine release elicits an inhibitory postsynaptic current in midbrain dopamine neurons. Neuron *42*, 939-946.

Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J., and Haussler, D. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature *441*, 87-90.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). Ultraconserved elements in the human genome. Science *304*, 1321-25.

Bellone, C., and Lüscher, C. (2005). mGluRs induce a long-term depression in the ventral tegmental area that involves a switch of the subunit composition of AMPA receptors. Eur J Neurosci *21*, 1280-88.

Benabid, A.L., Pollak, P., Gervason, C., Hoffmann, D., Gao, D.M., Hommel, M., Perret, J.E., and de Rougemont, J. (1991). Long-term suppression of tremor by chronic stimulation of the ventral intermediate thalamic nucleus. Lancet *337*, 403-06.

Bender, A., Krishnan, K.J., Morris, C.M., Taylor, G.A., Reeve, A.K., Perry, R.H., Jaros, E., Hersheson, J.S., Betts, J., et al. (2006). High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. Nat Genet *38*, 515-17.

Benner, E.J., Banerjee, R., Reynolds, A.D., Sherman, S., Pisarev, V.M., Tsiperson, V., Nemachek, C., Ciborowski, P., Przedborski, S., et al. (2008). Nitrated alpha-synuclein immunity accelerates degeneration of nigral dopaminergic neurons. PLoS ONE *3*, e1376.

Bennett, M.K., Miller, K.G., and Scheller, R.H. (1993). Casein kinase II phosphorylates the synaptic vesicle protein p65. J Neurosci *13*, 1701-07.

Bentivoglio, M., and Morelli, M. (2005). The organization and circuits of mesencephalic dopaminergic neurons and the distribution of dopamine receptors in the brain. Handbook of chemical neuroanatomy *21*, 1-106.

Bentivoglio, M., Bloom, F.E., Björklund, A., Dunnett, S.B., Hökfelt, T., Kaczmarek, L., Kuhar, M.J., Ottersen, O.P., Owman, C., et al. Handbook of chemical neuroanatomy (Elsevier).

Bentley DL. (2002) The mRNA assembly line: transcription and processing machines in the same factory. Curr Opin Cell Biol. *14*, 336-42.

Bentley DL. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol. *17*, 251-6.

Berg, D., Gerlach, M., Youdim, M.B., Double, K.L., Zecca, L., Riederer, P., and Becker, G. (2001). Brain iron pathways and their relevance to Parkinson's disease. J Neurochem *79*, 225-236.

Berger, B., Verney, C., Alvarez, C., Vigny, A., and Helle, K.B. (1985). New dopaminergic terminal fields in the motor, visual (area 18b) and retrosplenial cortex in the young and adult rat. Immunocytochemical and catecholamine histochemical analyses. Neuroscience *15*, 983-998.

Berget, S.M. (1995). Exon recognition in vertebrate splicing. J Biol Chem *270*, 2411-14.

Berman, S.B., and Hastings, T.G. (1997). Inhibition of glutamate transport in synaptosomes by dopamine oxidation and reactive oxygen species. J Neurochem *69*, 1185-195.

Bernheimer, H., Birkmayer, W., Hornykiewicz, O., Jellinger, K., and Seitelberger, F. (1973). Brain dopamine and the syndromes of Parkinson and Huntington. Clinical, morphological and neurochemical correlations. J Neurol Sci *20*, 415-455.

BERTLER, A., and ROSENGREN, E. (1959). Occurrence and distribution of dopamine in brain and other tissues. Experientia *15*, 10-11.

Bertoncini, C.W., Jung, Y.S., Fernandez, C.O., Hoyer, W., Griesinger, C., Jovin, T.M., and Zweckstetter, M. (2005). Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. Proc Natl Acad Sci U S A *102*, 1430-35.

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., et al. (2004). Global identification of human transcribed sequences with genome

tiling arrays. Science *306*, 2242-46.

Betarbet, R., Turner, R., Chockkan, V., DeLong, M.R., Allers, K.A., Walters, J., Levey, A.I., and Greenamyre, J.T. (1997). Dopaminergic neurons intrinsic to the primate striatum. J Neurosci *17*, 6761-68.

Beyer, K. (2006). Alpha-synuclein structure, posttranslational modification and alternative splicing as aggregation enhancers. Acta Neuropathol *112*, 237-251.

Beyer, K., Domingo-Sábat, M., Lao, J.I., Carrato, C., Ferrer, I., and Ariza, A. (2008). Identification and characterization of a new alpha-synuclein isoform and its role in Lewy body diseases. Neurogenetics *9*, 15-23.

Beyer, K., Humbert, J., Ferrer, A., Lao, J.I., Carrato, C., López, D., Ferrer, I., and Ariza, A. (2006). Low alpha-synuclein 126 mRNA levels in dementia with Lewy bodies and Alzheimer disease. Neuroreport *17*, 1327-330.

Beyer, K., Lao, J.I., Carrato, C., Mate, J.L., López, D., Ferrer, I., and Ariza, A. (2004). Differential expression of alpha-synuclein isoforms in dementia with Lewy bodies. Neuropathol Appl Neurobiol *30*, 601-07.

Bhinge AA, Kim J, Euskirchen GM, Snyder M, Iyer VR. (2007). Mapping the chromosomal targets of STAT1 by Sequence Tag Analysis of GenomicEnrichment (STAGE). Genome Res. *17*, 910-16.

Biere, A.L., Wood, S.J., Wypych, J., Steavenson, S., Jiang, Y., Anafi, D., Jacobsen, F.W., Jarosinski, M.A., Wu, G.M., et al. (2000). Parkinson's disease-associated alpha-synuclein is more fibrillogenic than beta- and gamma-synuclein and cannot cross-seed its homologs. J Biol Chem *275*, 34574-79.

Birnstiel, M.L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3' processing: the end is in site!. Cell *41*, 349-359.

Bisaglia, M., Schievano, E., Caporale, A., Peggion, E., and Mammi, S. (2006). The 11-mer repeats of human alpha-synuclein in vesicle interactions and lipid composition discrimination: a cooperative role. Biopolymers *84*, 310-16.

Bisaglia, M., Trolio, A., Bellanda, M., Bergantino, E., Bubacco, L., and Mammi, S. (2006). Structure and topology of the non-amyloid-beta component fragment of human alpha- synuclein bound to micelles: implications for the aggregation process. Protein Sci *15*, 1408-416.

Bishop JO, Morton JG, Rosbash M, Richardson M. (1974). Three abundance classes in HeLa cell messenger RNA. Nature *250*, 199-204.

Bjarkam, C.R., and Sørensen, J.C. (2004). Therapeutic strategies for neurodegenerative disorders: emerging clues from Parkinson's disease. Biol Psychiatry *56*, 213-16.

Björklund, A., and Dunnett, S.B. (2007). Dopamine neuron systems in the brain: an update. Trends Neurosci *30*, 194-202.

Björklund, A., and Dunnett, S.B. (2007). Fifty years of dopamine research. Trends Neurosci *30*, 185-87.

Björklund, A., and Nobin, A. (1973). Fluorescence histochemical and microspectrofluorometric mapping of dopamine and noradrenaline cell groups in the rat diencephalon. Brain Res *51*, 193-205.

Björklund, A., and Stenevi, U. (1979). Reconstruction of the nigrostriatal dopamine pathway by intracerebral nigral transplants. Brain Res *177*,

555-560.

Blandini, F., Nappi, G., Tassorelli, C., and Martignoni, E. (2000). Functional changes of the basal ganglia circuitry in Parkinson's disease. Prog Neurobiol *62*, 63-88.

Blaschke RJ, Töpfer C, Marchini A, Steinbeisser H, Janssen JW, Rappold GA. (2003). Transcriptional and translational regulation of the Leri-Weill and Turner syndrome homeobox gene SHOX. J Biol Chem *278*, 47820-6.

Bonaldo MF, Lennon G, Soares MB. (1996). Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res. *6*, 791-806.

Bonci, A., and Malenka, R.C. (1999). Properties and plasticity of excitatory synapses on dopaminergic and GABAergic cells in the ventral tegmental area. J Neurosci *19*, 3723-730.

Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell *123*, 1279-91.

Bossy-Wetzel, E., Barsoum, M.J., Godzik, A., Schwarzenbacher, R., and Lipton, S.A. (2003). Mitochondrial fission in apoptosis, neurodegeneration and aging. Curr Opin Cell Biol *15*, 706-716.

Bourdon JC, Fernandes K, Murray-Zmijewski F, Liu G, Diot A, Xirodimas DP, Saville MK, Lane DP. (2005). p53 isoforms can regulate p53 transcriptional activity. Genes Dev *19*, 2122-37.

Braak, H., and Braak, E. (2000). Pathoanatomy of Parkinson's disease. J Neurol *247 Suppl 2*, II3-I10.

Braak, H., Bohl, J.R., Müller, C.M., Rüb, U., de Vos, R.A., and Del Tredici, K. (2006). Stanley Fahn Lecture 2005: The staging procedure for the inclusion body pathology associated with sporadic Parkinson's disease reconsidered. Mov Disord *21*, 2042-051.

Braak, H., Rüb, U., and Del Tredici, K. (2006). Cognitive decline correlates with neuropathological stage in Parkinson's disease. J Neurol Sci *248*, 255-58.

Braak, H., Tredici, K.D., Gai, W.P., and Braak, E. (2000). Alpha-synuclein is not a requisite component of synaptic boutons in the adult human central nervous system. Journal of Chemical Neuroanatomy *20*, 245-252.

Britten, R.J., and Davidson, E.H. (1969). Gene regulation for higher cells: a theory. Science *165*, 349-357.

Brockdorff, N. (1998). The role of Xist in X-inactivation. Curr Opin Genet Dev *8*, 328-333.

Brodsky AS, Meyer CA, Swinburne IA, Hall G, Keenan BJ, Liu XS, Fox EA, Silver PA. (2005). Genomic mapping of RNA polymerase II reveals sites of co-transcriptionalregulation in human cells. Genome Biol. *6*, R64.

Brundin P., & Björklund A. (1998). Nat Neurosci. In *Survival of expanded dopaminergic precursors is critical for clinical trials..* USA

Burns, R.S., Chiueh, C.C., Markey, S.P., Ebert, M.H., Jacobowitz, D.M., and Kopin, I.J. (1983). A primate model of parkinsonism: selective destruction of dopaminergic neurons in the pars compacta of the substantia nigra by N-methyl-4-phenyl-1,2,3,6-tetrahydropyridine. Proc Natl Acad Sci U S A *80*, 4546-550.

Bussell, R., and Eliezer, D. (2004). Effects of Parkinson's disease-linked mutations on the structure of lipid-associated alpha-synuclein.

Biochemistry *43*, 4810-18.

Bussell, R., Ramlall, T.F., and Eliezer, D. (2005). Helix periodicity, topology, and dynamics of membrane-associated alpha-synuclein. Protein Sci *14*, 862-872.

Campion, D., Martin, C., Heilig, R., Charbonnier, F., Moreau, V., Flaman, J.M., Petit, J.L., Hannequin, D., Brice, A., and Frebourg, T. (1995). The NACP/synuclein gene: chromosomal assignment and screening for alterations in Alzheimer disease. Genomics *26*, 254-57.

CARLSSON, A., LINDQVIST, M., MAGNUSSON, T., and WALDECK, B. (1958). On the presence of 3-hydroxytyramine in brain. Science *127*, 471.

Carmichael G. G. (2003). *Antisense starts making more sense*. Nat Biotechnol. 21(4), 371-2

Carneiro, A.M., Ingram, S.L., Beaulieu, J.M., Sweeney, A., Amara, S.G., Thomas, S.M., Caron, M.G., and Torres, G.E. (2002). The multiple LIM domain-containing adaptor protein Hic-5 synaptically colocalizes and interacts with the dopamine transporter. J Neurosci *22*, 7045-054.

Carninci, P. (2007). Constructing the landscape of the mammalian transcriptome. J Exp Biol *210*, 1497-1506.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., et al. (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559-563.

Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki, N., et al. (1996). High-efficiency full-length cDNA cloning by biotinylated CAP trapper. Genomics *37*, 327-336.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engström, P.G., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Carninci P, Shibata Y, Hayatsu N, Itoh M, Shiraki T, Hirozane T, Watahiki A,Shibata K, Konno H, Muramatsu M, Hayashizaki Y. (2001). Balanced-size and long-size cloning of full-length, cap-trapped cDNAs intovectors of the novel lambda-FLC family allows enhanced gene discovery rate and functional analysis. Genomics 77, 79-90.

Carninci P, Waki K, Shiraki T, Konno H, Shibata K, Itoh M, Aizawa K, Arakawa T,Ishii Y, Sasaki D, Bono H, Kondo S, Sugahara Y, Saito R, Osato N, Fukuda S, Sato K, Watahiki A, Hirozane-Kishikawa T, Nakamura M, Shibata Y, Yasunishi A, Kikuchi N, Yoshiki A, Kusakabe M, Gustincich S, Beisel K, Pavan W, Aidinis V, Nakagawara A, Held WA, Iwata H, Kono T, Nakauchi H, Lyons P, Wells C, Hume DA, Fagiolini M, Hensch TK, Brinkmeier M, Camper S, Hirota J, Mombaerts P, Muramatsu M, Okazaki Y,Kawai J, Hayashizaki Y. (2003). Targeting a complex transcriptome: the construction of the mouse full-length cDNAencyclopedia. Genome Res. 13, 1273-89.

Carninci, P., Westover, A., Nishiyama, Y., Ohsumi, T., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., et al. (1997). High efficiency selection of full-length cDNA by improved biotinylated cap trapper. DNA Res *4*, 61-66.

Carter NP. (1994). Cytogenetic analysis by chromosome painting. Cytometry *18*, 2-10.

Cassarino, D.S., and Bennett, J.P. (1999). An evaluation of the role of mitochondria in neurodegenerative diseases: mitochondrial mutations and oxidative pathology, protective nuclear responses, and cell death in neurodegeneration. Brain Res Brain Res Rev *29*, 1-25.

Castelo-Branco, G., Wagner, J., Rodriguez, F.J., Kele, J., Sousa, K., Rawal, N., Pasolli, H.A., Fuchs, E., Kitajewski, J., and Arenas, E. (2003). Differential regulation of midbrain dopaminergic neuron development by Wnt-1, Wnt-3a, and Wnt-5a. Proc Natl Acad Sci U S A *100*, 12747-752.

Castner, S.A., and Goldman-Rakic, P.S. (2004). Enhancement of working memory in aged monkeys by a sensitizing regimen of dopamine D1 receptor stimulation. J Neurosci *24*, 1446-450.

Castner, S.A., Goldman-Rakic, P.S., and Williams, G.V. (2004). Animal models of working memory: insights for targeting cognitive dysfunction in schizophrenia. Psychopharmacology (Berl) *174*, 111-125.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499-509.

Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ,Morris MS, Fodor SP. (1996). Accessing genetic information with high-density DNA arrays. Science *274*, 610-4.

Chen, J., Sun, M., Lee, S., Zhou, G., Rowley, J.D., and Wang, S.M. (2002). Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. Proc Natl Acad Sci U S A *99*, 12257-262.

Chen, L., and Feany, M.B. (2005). Alpha-synuclein phosphorylation controls neurotoxicity and inclusion formation in a Drosophila model of Parkinson disease. Nat Neurosci *8*, 657-663.

Chen J., Sun M., Kent W.J., Hunag X., Xie H., Wang w., Zhou G., Shi R.Z., Rowley J.D. (2004) Over 20% of human transcripts might form sense-antisense pairs. Nucleic Acids Res. *32*, 4812-20.

Chen J., Sun M., Hurst L.D., Carmichael G.G., Rowley J.D. (2005). Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. Trends Genet. *21*, 326-29.

Chen T, Ueda Y, Xie S, Li E. (2002). A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. J Biol Chem *277*, 38746-54.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., et al. (2005). Transcriptional maps of 10 human chromosomes at 5- nucleotide resolution. Science *308*, 1149-154.

Chow, N., Cox, C., Callahan, L.M., Weimer, J.M., Guo, L., and Coleman, P.D. (1998). Expression profiles of multiple genes in single neurons of Alzheimer's disease. Proc Natl Acad Sci U S A *95*, 9620-25.

Chu, Y., and Kordower, J.H. (2007). Age-associated increases of alpha-synuclein in monkeys and humans are associated with nigrostriatal

dopamine depletion: Is this the target for Parkinson's disease? Neurobiol Dis *25*, 134-149.

Chung, C.Y., Seo, H., Sonntag, K.C., Brooks, A., Lin, L., and Isacson, O. (2005). Cell type- specific gene expression of midbrain dopaminergic neurons reveals molecules involved in their vulnerability and protection. Hum Mol Genet *14*, 1709-725.

Chung, S., Hedlund, E., Hwang, M., Kim, D.W., Shin, B.S., Hwang, D.Y., Jung Kang, U., Isacson, O., and Kim, K.S. (2005). The homeodomain transcription factor Pitx3 facilitates differentiation of mouse embryonic stem cells into AHD2-expressing dopaminergic neurons. Mol Cell Neurosci *28*, 241-252.

Clark I. E., Dodson M. W., Jiang C., Cao J. H., Huh J. R., et al. (2006). In *Drosophila pink1 is required for mitochondrial function and interacts genetically with parkin.*.Nature 441, 1162-1166

Clayton, D.A. (1991). Replication and transcription of vertebrate mitochondrial DNA. Annu Rev Cell Biol *7*, 453-478.

Clayton, D.F., and George, J.M. (1998). The synucleins: a family of proteins involved in synaptic function, plasticity, neurodegeneration and disease. Trends Neurosci *21*, 249-254.

Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., et al. (2008). Stem cell transcriptome profiling via massive- scale mRNA sequencing. Nat Methods *5*, 613-19.

Clotman, F., Van Maele-Fabry, G., and Picard, J.J. (1997). Retinoic acid induces a tissue- specific deletion in the expression domain of Otx2. Neurotoxicol Teratol *19*, 163-69.

Cole, N.B., Dieuliis, D., Leo, P., Mitchell, D.C., and Nussbaum, R.L. (2008). Mitochondrial translocation of alpha-synuclein is promoted by intracellular acidification. Exp Cell Res *314*, 2076-089.

Conway, K.A., Harper, J.D., and Lansbury, P.T. (2000). Fibrils formed in vitro from alpha-synuclein and two mutant forms linked to Parkinson's disease are typical amyloid. Biochemistry *39*, 2552-563.

Conway, K.A., Rochet, J.C., Bieganski, R.M., and Lansbury, P.T. (2001). Kinetic stabilization of the alpha-synuclein protofibril by a dopamine-alpha-synuclein adduct. Science *294*, 1346-49.

Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM. (2006). Comprehensive analysis of transcriptional promoter structure and function in 1%of the human genome. Genome Res *16*, 1-10.

Cotzias, G.C., Van Woert, M.H., and Schiffer, L.M. (1967). Aromatic amino acids and modification of parkinsonism. N Engl J Med *276*, 374-79.

Coupar BE, Davies JA, Chesterton CJ. (1978). Quantification of hepatic transcribing RNA polymerase molecules,polyribonucleotide elongation rates and messenger RNA complexity in fed and fasted rats. Eur J Biochem. *84*, 611-23.

Courtois V, Chatelain G, Han ZY, Le Novère N, Brun G, Lamonerie T. (2003). New Otx2 mRNA isoforms expressed in the mouse brain. J Neurochem *84*, 840-53.

Crabtree G.R., Schibler U., and Scott M.P. (1992). Transcriptional regulatory mechanisms in liver and midgut morphogenesis of vertebrates and invertebrates. In Transcriptional Regualtion, (McKnight S.L., and Yamamoto K.R., eds) pp. 1063-1102, Cold Spring Harbor Laboratory Press, New York.

Crino, P., Khodakhah, K., Becker, K., Ginsberg, S., Hemby, S., and Eberwine, J. (1998). Presence and phosphorylation of transcription factors in developing dendrites. Proc Natl Acad Sci U S A *95*, 2313-18.

Crino, P.B., Trojanowski, J.Q., Dichter, M.A., and Eberwine, J. (1996). Embryonic neuronal markers in tuberous sclerosis: single-cell molecular pathology. Proc Natl Acad Sci U S A *93*, 14152-57.

Crossley, P.H., Martinez, S., and Martin, G.R. (1996). Midbrain development induced by FGF8 in the chick embryo. Nature *380*, 66-68.

Cwirla SE, Peters EA, Barrett RW, Dower WJ. (1990). Peptides on phage: a vast library of peptides for identifying ligands. Proc Natl Acad Sci U S A. *87*, 6378-82.

Dahlström, A., and Fuxe, K. (1964). Localization of monoamines in the lower brain stem. Experientia *20*, 398-99.

Dammann R, Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP. (2000). Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. Nat Genet *25*, 315-9.

Danielian, P.S., and McMahon, A.P. (1996). Engrailed-1 as a target of the Wnt-1 signalling pathway in vertebrate midbrain development. Nature *383*, 332-34.

Daniels, G.M., and Amara, S.G. (1999). Regulated trafficking of the human dopamine transporter. Clathrin-mediated internalization and lysosomal degradation in response to phorbol esters. J Biol Chem *274*, 35794-5801.

Das M., Harvey I., Chu L.L., Sinha M., Pelletier J. (2001). Full-length cDNAs: more than just reaching the ends. Physiol Genomics *6*, 57-80.

Dauer, W., and Przedborski, S. (2003). Parkinson's disease: mechanisms and models. Neuron *39*, 889-909.

Davidson, W.S., Jonas, A., Clayton, D.F., and George, J.M. (1998). Stabilization of alpha- synuclein secondary structure upon binding to synthetic membranes. J Biol Chem *273*, 9443-49.

Davis, C.A., and Ares, M. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *103*, 3262-67.

Davuluri, R.V., Suzuki, Y., Sugano, S., Plass, C., and Huang, T.H. (2008). The functional consequences of alternative promoter use in mammalian genomes. Trends Genet *24*, 167-177.

Daws, L.C., Callaghan, P.D., Morón, J.A., Kahlig, K.M., Shippenberg, T.S., Javitch, J.A., and Galli, A. (2002). Cocaine increases dopamine uptake and cell surface expression of dopamine transporters. Biochem Biophys Res Commun *290*, 1545-550.

Dearry, A., Gingrich, J.A., Falardeau, P., Fremeau, R.T., Bates, M.D., and Caron, M.G. (1990). Molecular cloning and expression of the gene for a human D1 dopamine receptor. Nature *347*, 72-76.

DeBlock, M., and Debrouwer, D. (1996). RNA-RNA in situ hybridization using DIG-labeled probes: the effect of high molecular weight polyvinyl alcohol on the alkaline phosphatase indoxyl-nitroblue tetrazolium reaction. Nonradioactive in situ

hybridization application manual, 2nd ed. Boehringer Mannheim Biochemicals, Indianapolis, Ind , 141-45.

DeCerbo J, Carmichael GG. (2005). Retention and repression: fates of hyperedited RNAs in the nucleus. Curr Opin Cell Biol *17*, 302-8.

De Hoon M. and Hayashizaki Y. (2008). Deep cap analysis gene expression (CAGE): genome-wide identification ofpromoters, quantification of their expression, and network inference. Biotechniques *44*, 627-8, 630, 632.

Dekker J, Rippe K, Dekker M, Kleckner N. (2002). Capturing chromosome conformation. Science *295*, 1306-11.

Denoeud, F., Kapranov, P., Ucla, C., Frankish, A., Castelo, R., Drenkow, J., Lagarde, J., Alioto, T., Manzano, C., et al. (2007). Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. Genome Res *17*, 746-759.

Dermitzakis, E.T., Reymond, A., Lyle, R., Scamuffa, N., Ucla, C., Deutsch, S., Stevenson, B.J., Flegel, V., Bucher, P., et al. (2002). Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature *420*, 578-582.

Dickson, B.J., and Gilestro, G.F. (2006). Regulation of commissural axon pathfinding by slit and its Robo receptors. Annu Rev Cell Dev Biol *22*, 651-675.

Doss-Pepe, E.W., Chen, L., and Madura, K. (2005). Alpha-synuclein and parkin contribute to the assembly of ubiquitin lysine 63-linked multiubiquitin chains. J Biol Chem *280*, 16619-624.

Draghici S, Khatri P, Eklund AC, Szallasi Z. (2006). Reliability and reproducibility issues in DNA microarray measurements. Trends Genet. *22*, 101-9.

Drolet, R.E., Behrouz, B., Lookingland, K.J., and Goudreau, J.L. (2006). Substrate-mediated enhancement of phosphorylated tyrosine hydroxylase in nigrostriatal dopamine neurons: evidence for a role of alpha-synuclein. J Neurochem *96*, 950-59.

Dye MJ, Gromak N, Proudfoot NJ. (2006). Exon tethering in transcription by RNA polymerase II. Mol Cell *21*, 849-59.

Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M., and Coleman, P. (1992). Analysis of gene expression in single live neurons. Proc Natl Acad Sci U S A *89*, 3010-14.

Echelard, Y., Epstein, D.J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J.A., and McMahon, A.P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell *75*, 1417-430.

Eddy, S.R. (1999). Noncoding RNA genes. Curr Opin Genet Dev *9*, 695-99.

Eliezer, D., Kutluay, E., Bussell, R., and Browne, G. (2001). Conformational properties of alpha-synuclein in its free and lipid-associated states. J Mol Biol *307*, 1061-073.

ENCODE Project Consortium, Birney, E.Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Erdmann, V.A., Szymanski, M., Hochberg, A., de Groot, N., and Barciszewski, J. (1999). Collection of mRNA-like non-coding RNAs. Nucleic Acids Res *27*, 192-95.

Erickson, J.D., Eiden, L.E., and Hoffman, B.J. (1992). Expression cloning of a reserpine- sensitive vesicular monoamine transporter. Proc Natl Acad Sci U S A *89*, 10993-97.

Erickson, J.D., Schafer, M.K., Bonner, T.I., Eiden, L.E., and Weihe, E. (1996). Distinct pharmacological properties and distribution in neurons and endocrine cells of two isoforms of the human vesicular monoamine transporter. Proc Natl Acad Sci U S A *93*, 5166-171.

Exner, N., Treske, B., Paquet, D., Holmström, K., Schiesling, C., Gispert, S., Carballo- Carbajal, I., Berg, D., Hoepken, H.H., et al. (2007). Loss-of-function of human PINK1 results in mitochondrial pathology and can be rescued by parkin. J Neurosci *27*, 12413-18.

Fahey ME, Moore TF, Higgins DG. (2002). Overlapping antisense transcription in the human genome. Comp Funct Genomics *3*, 244-53.

Fahn, S. (2003). Description of Parkinson's disease as a clinical syndrome. Ann N Y Acad Sci *991*, 1-14.

Fairbrother, W.G., Yeh, R.F., Sharp, P.A., and Burge, C.B. (2002). Predictive identification of exonic splicing enhancers in human genes. Science *297*, 1007-013.

Falck, B., and Hillarp, N. Thieme, G., Torp, A. 1962. Fluorescence of catecholamines and related compounds condensed with formaldehyde. J. Histochem. Cytochem *10*, 348-354.

Farkas, L.M., Dünker, N., Roussa, E., Unsicker, K., and Krieglstein, K. (2003). Transforming growth factor-beta(s) are essential for the development of midbrain dopaminergic neurons in vitro and in vivo. J Neurosci *23*, 5178-186.

Faulkner, G.J., Forrest, A.R., Chalk, A.M., Schroder, K., Hayashizaki, Y., Carninci, P., Hume, D.A., and Grimmond, S.M. (2008). A rescue strategy for multimapping short sequence tags refines surveys of transcriptional activity by CAGE. Genomics *91*, 281-88.

Fazeli, A., Dickinson, S.L., Hermiston, M.L., Tighe, R.V., Steen, R.G., Small, C.G., Stoeckli, E.T., Keino-Masu, K., Masu, M., et al. (1997). Phenotype of mice lacking functional Deleted in colorectal cancer (Dcc) gene. Nature *386*, 796-804.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-46.

Flores, C., Manitt, C., Rodaros, D., Thompson, K.M., Rajabi, H., Luk, K.C., Tritsch, N.X., Sadikot, A.F., Stewart, J., and Kennedy, T.E. (2005). Netrin receptor deficient mice exhibit functional reorganization of dopaminergic systems and do not sensitize to amphetamine. Mol Psychiatry *10*, 606-612.

Forno, L.S., DeLanney, L.E., Irwin, I., and Langston, J.W. (1996). Electron microscopy of Lewy bodies in the amygdala-parahippocampal region. Comparison with inclusion bodies in the MPTP-treated squirrel monkey. Adv Neurol *69*, 217-228.

Fortin, D.L., Nemani, V.M., Voglmaier, S.M., Anthony, M.D., Ryan, T.A., and Edwards, R.H. (2005). Neural activity controls the synaptic accumulation of alpha-synuclein. J Neurosci *25*,

242

10913-921.

Frank, S. (2006). Dysregulation of mitochondrial fusion and fission: an emerging concept in neurodegeneration. Acta Neuropathol *111*, 93-100.

Freed, C.R., Greene, P.E., Breeze, R.E., Tsai, W.Y., DuMouchel, W., Kao, R., Dillon, S., Winfield, H., Culver, S., et al. (2001). Transplantation of embryonic dopamine neurons for severe Parkinson's disease. N Engl J Med *344*, 710-19.

Freed, W.J., Perlow, M.J., Karoum, F., Seiger, A., Olson, L., Hoffer, B.J., and Wyatt, R.J. (1980). Restoration of dopaminergic function by grafting of fetal rat substantia nigra to the caudate nucleus: long-term behavioral, biochemical, and histochemical studies. Ann Neurol *8*, 510-19.

Freeman, T.B., Spence, M.S., Boss, B.D., Spector, D.H., Strecker, R.E., Olanow, C.W., and Kordower, J.H. (1991). Development of dopaminergic neurons in the human substantia nigra. Exp Neurol *113*, 344-353.

Freund, T.F., Bolam, J.P., Björklund, A., Stenevi, U., Dunnett, S.B., Powell, J.F., and Smith, A.D. (1985). Efferent synaptic connections of grafted dopaminergic neurons reinnervating the host neostriatum: a tyrosine hydroxylase immunocytochemical study. J Neurosci *5*, 603-616.

Frith, M.C., Valen, E., Krogh, A., Hayashizaki, Y., Carninci, P., and Sandelin, A. (2008). A code for transcription initiation in mammalian genomes. Genome Res *18*, 1-12.

Fujiwara, H., Hasegawa, M., Dohmae, N., Kawashima, A., Masliah, E., Goldberg, M.S., Shen, J., Takio, K., and Iwatsubo, T. (2002). alpha-Synuclein is phosphorylated in synucleinopathy lesions. Nat Cell Biol *4*, 160-64.

Fukagawa T., Nogami M., Yoshikawa M., Ikeno M., Okazaki T., Takami Y., Nakayama T., Oshimura M. (2004). Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nat Cell Biol. *6*, 784-91.

Funayama, M., Hasegawa, K., Kowa, H., Saito, M., Tsuji, S., and Obata, F. (2002). A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. Ann Neurol *51*, 296-301.

Gainetdinov, R.R., Fumagalli, F., Jones, S.R., and Caron, M.G. (1997). Dopamine transporter is required for in vivo MPTP neurotoxicity: evidence from mice lacking the transporter. J Neurochem *69*, 1322-25.

Gainetdinov, R.R., Fumagalli, F., Wang, Y.M., Jones, S.R., Levey, A.I., Miller, G.W., and Caron, M.G. (1998). Increased MPTP neurotoxicity in vesicular monoamine transporter 2 heterozygote knockout mice. J Neurochem *70*, 1973-78.

Galpern, W.R., and Lang, A.E. (2006). Interface between tauopathies and synucleinopathies: a tale of two proteins. Ann Neurol *59*, 449-458.

Galvin, J.E. (2006). Interaction of alpha-synuclein and dopamine metabolites in the pathogenesis of Parkinson's disease: a case for the selective vulnerability of the substantia nigra. Acta Neuropathol *112*, 115-126.

Garris, P.A., Collins, L.B., Jones, S.R., and Wightman, R.M. (1993). Evoked extracellular dopamine in vivo in the medial prefrontal cortex. J Neurochem *61*, 637-647.

Gasser, T. (2007). Update on the genetics of Parkinson's disease. Mov Disord *22 Suppl 17*, S343-350.

Gates, M.A., Torres, E.M., White, A., Fricker-Gates, R.A., and Dunnett, S.B. (2006). Re- examining the ontogeny of substantia nigra dopamine neurons. Eur J Neurosci *23*, 1384-390.

Gauthier, J., Parent, M., Lévesque, M., and Parent, A. (1999). The axonal arborization of single nigrostriatal neurons in rats. Brain Res *834*, 228-232.

George, J.M., Jin, H., Woods, W.S., and Clayton, D.F. (1995). Characterization of a novel protein regulated during the critical period for song learning in the zebra finch. Neuron *15*, 361-372.

Gerfen, C.R., and Wilson, C.J. (1996). The basal ganglia. Handbook of chemical neuroanatomy *12*, 371-468.

Gerfen, C.R., Baimbridge, K.G., and Thibault, J. (1987). The neostriatal mosaic: III. Biochemical and developmental dissociation of patch-matrix mesostriatal systems. J Neurosci *7*, 3935-944.

Gerfen, C.R., Herkenham, M., and Thibault, J. (1987). The neostriatal mosaic: II. Patch- and matrix-directed mesostriatal dopaminergic and non-dopaminergic systems. J Neurosci *7*, 3915-934.

German, D.C., and Liang, C.L. (1993). Neuroactive peptides exist in the midbrain dopaminergic neurons that contain calbindin-D28k. Neuroreport *4*, 491-94.

German, D.C., Manaye, K., Smith, W.K., Woodward, D.J., and Saper, C.B. (1989). Midbrain dopaminergic cell loss in Parkinson's disease: computer visualization. Ann Neurol *26*, 507-514.

German, D.C., Manaye, K.F., White, C.L., Woodward, D.J., McIntire, D.D., Smith, W.K., Kalaria, R.N., and Mann, D.M. (1992). Disease-specific patterns of locus coeruleus cell loss. Ann Neurol *32*, 667-676.

Giasson, B.I., Uryu, K., Trojanowski, J.Q., and Lee, V.M. (1999). Mutant and wild type human alpha-synucleins assemble into elongated filaments with distinct morphologies in vitro. J Biol Chem *274*, 7619-622.

Gibb, W.R., and Lees, A.J. (1988). The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. J Neurol Neurosurg Psychiatry *51*, 745-752.

Gilbert, W. (1978). Why genes in pieces? Nature *271*, 501.

Gill, S.S., Patel, N.K., Hotton, G.R., O'Sullivan, K., McCarter, R., Bunnage, M., Brooks, D.J., Svendsen, C.N., and Heywood, P. (2003). Direct brain infusion of glial cell line-derived neurotrophic factor in Parkinson disease. Nat Med *9*, 589-595.

Giros, B., Jaber, M., Jones, S.R., Wightman, R.M., and Caron, M.G. (1996). Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. Nature *379*, 606-612.

Giros, B., Wang, Y.M., Suter, S., McLeskey, S.B., Pifl, C., and Caron, M.G. (1994). Delineation of discrete domains for substrate, cocaine, and tricyclic antidepressant interactions using chimeric dopamine-norepinephrine transporters. J Biol Chem *269*, 15985-88.

Goff, L.A., Bowers, J., Schwalm, J., Howerton, K., Getts, R.C., and Hart, R.P. (2004). Evaluation of sense-strand mRNA amplification by

comparative quantitative PCR. BMC Genomics *5*, 76.

Gogos, J.A., Morgan, M., Luine, V., Santha, M., Ogawa, S., Pfaff, D., and Karayiorgou, M. (1998). Catechol-O-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. Proc Natl Acad Sci U S A *95*, 9991-96.

Golbe, L.I. (1991). Young-onset Parkinson's disease: a clinical review. Neurology *41*, 168-173.

Goldberg, R.B., Hoschek, G., Kamalay, J.C., and Timberlake, W.E. (1978). Sequence complexity of nuclear and polysomal RNA in leaves of the tobacco plant. Cell *14*, 123-131.

Goldman-Rakic, P.S., Castner, S.A., Svensson, T.H., Siever, L.J., and Williams, G.V. (2004). Targeting the dopamine D1 receptor in schizophrenia: insights for cognitive dysfunction. Psychopharmacology (Berl) *174*, 3-16.

Gomez-Lazaro, M., Bonekamp, N.A., Galindo, M.F., Jordán, J., and Schrader, M. (2008). 6-Hydroxydopamine (6-OHDA) induces Drp1-dependent mitochondrial fragmentation in SH-SY5Y cells. Free Radic Biol Med *44*, 1960-69.

Gonzalez, A.M., Walther, D., Pazos, A., and Uhl, G.R. (1994). Synaptic vesicular monoamine transporter expression: distribution and pharmacologic profile. Brain Res Mol Brain Res *22*, 219-226.

Gorbatyuk, O.S., Li, S., Sullivan, L.F., Chen, W., Kondrikova, G., Manfredsson, F.P., Mandel, R.J., and Muzyczka, N. (2008). The phosphorylation state of Ser-129 in human alpha- synuclein determines neurodegeneration in a rat model of Parkinson disease. Proc Natl Acad Sci U S A *105*, 763-68.

Granérus, A.K. (1978). Factors influencing the occurrence of "on-off" symptoms during long-term treatment with L-dopa. Acta Med Scand *203*, 75-85.

Graveley, B.R. (2001). Alternative splicing: increasing diversity in the proteomic world. Trends Genet *17*, 100-07.

Graybiel, A.M., and Ragsdale, C.W. (1983). Biochemical anatomy of the striatum. Chemical Neuroanatomy , 427-504.

Greenamyre, J.T., Betarbet, R., and Sherer, T.B. (2003). The rotenone model of Parkinson's disease: genes, environment and mitochondria. Parkinsonism Relat Disord *9 Suppl 2*, S59- S64.

Greenbaum, E.A., Graves, C.L., Mishizen-Eberz, A.J., Lupoli, M.A., Lynch, D.R., Englander, S.W., Axelsen, P.H., and Giasson, B.I. (2005). The E46K mutation in alpha-synuclein increases amyloid fibril formation. J Biol Chem *280*, 7800-07.

Greene, J.G., Dingledine, R., and Greenamyre, J.T. (2005). Gene expression profiling of rat midbrain dopamine neurons: implications for selective vulnerability in parkinsonism. Neurobiol Dis *18*, 19-31.

Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., and Fraser, P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. Mol Cell *5*, 377-386.

Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res *32*, D109-111.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res *34*, D140-44.

Grillner, P., and Mercuri, N.B. (2002). Intrinsic membrane properties and synaptic inputs regulating the firing activity of the dopamine neurons. Behav Brain Res *130*, 149-169.

Grimm, J., Mueller, A., Hefti, F., and Rosenthal, A. (2004). Molecular basis for catecholaminergic neuron diversity. Proc Natl Acad Sci U S A *101*, 13891-96.

Gross, S.D., Hoffman, D.P., Fisette, P.L., Baas, P., and Anderson, R.A. (1995). A phosphatidylinositol 4,5-bisphosphate-sensitive casein kinase I alpha associates with synaptic vesicles and phosphorylates a subset of vesicle proteins. J Cell Biol *130*, 711-724.

Grossman, M.H., Creveling, C.R., Rybczynski, R., Braverman, M., Isersky, C., and Breakefield, X.O. (1985). Soluble and particulate forms of rat catechol-O-methyltransferase distinguished by gel electrophoresis and immune fixation. J Neurochem *44*, 421-432.

Grzybowska, E.A., Wilczynska, A., and Siedlecki, J.A. (2001). Regulatory functions of 3'UTRs. Biochem Biophys Res Commun *288*, 291-95.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Guillin, O., Diaz, J., Carroll, P., Griffon, N., Schwartz, J.C., and Sokoloff, P. (2001). BDNF controls dopamine D3 receptor expression and triggers behavioural sensitization. Nature *411*, 86-89.

Guldberg, H.C., and Marsden, C.A. (1975). Catechol-O-methyl transferase: pharmacological aspects and physiological role. Pharmacol Rev *27*, 135-206.

Gustincich, S., Contini, M., Gariboldi, M., Puopolo, M., Kadota, K., Bono, H., LeMieux, J., Walsh, P., Carninci, P., et al. (2004). Gene discovery in genetically labeled single dopaminergic neurons of the retina. Proc Natl Acad Sci U S A *101*, 5069-074.

Gustincich, S., Sandelin, A., Plessy, C., Katayama, S., Simone, R., Lazarevic, D., Hayashizaki, Y., and Carninci, P. (2006). The complexity of the mammalian transcriptome. J Physiol *575*, 321-332.

Hagell, P., Piccini, P., Björklund, A., Brundin, P., Rehncrona, S., Widner, H., Crabb, L., Pavese, N., Oertel, W.H., et al. (2002). Dyskinesias following neural transplantation in Parkinson's disease. Nat Neurosci *5*, 627-28.

Hald, A., and Lotharius, J. (2005). Oxidative stress and inflammation in Parkinson's disease: is there a causal link? Exp Neurol *193*, 279-290.

Hallett, P.J., and Standaert, D.G. (2004). Rationale for and use of NMDA receptor antagonists in Parkinson's disease. Pharmacol Ther *102*, 155-174.

Halliday, G.M., and McCann, H. (2008). Human-based studies on alpha-synuclein deposition and relationship to Parkinson's disease symptoms. Exp Neurol *209*, 12-21.

Han, H., Weinreb, P.H., and Lansbury, P.T. (1995). The core Alzheimer's peptide NAC forms amyloid fibrils which seed and are seeded by beta-amyloid: is NAC a common trigger or target in neurodegenerative disease? Chem Biol *2*, 163-69.

Han J, Kim D, Morris KV. (2007). Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. Proc Natl Acad Sci U

S A *104*, 12422-7.

Hanada, K., Zhang, X., Borevitz, J.O., Li, W.H., and Shiu, S.H. (2007). A large number of novel coding small open reading frames in the intergenic regions of the Arabidopsis thaliana genome are transcribed and/or under purifying selection. Genome Res *17*, 632-640.

Haque, N.S., LeBlanc, C.J., and Isacson, O. (1997). Differential dissection of the rat E16 ventral mesencephalon and survival and reinnervation of the 6-OHDA-lesioned striatum by a subset of aldehyde dehydrogenase-positive TH neurons. Cell Transplant *6*, 239-248.

Harbers M., Carninci P. (2005). Tag-based approaches for transcriptome research and genome annotation. Nat Methods *2*, 495-502.

Hardy, J., Cookson, M.R., and Singleton, A. (2003). Genes and parkinsonism. Lancet Neurol *2*, 221-28.

Harris, G.C., Wimmer, M., Byrne, R., and Aston-Jones, G. (2004). Glutamate-associated plasticity in the ventral tegmental area is necessary for conditioning environmental stimuli with morphine. Neuroscience *129*, 841-47.

Harrower, T.P., Michell, A.W., and Barker, R.A. (2005). Lewy bodies in Parkinson's disease: protectors or perpetrators? Exp Neurol *195*, 1-6.

Hastrup, H., Karlin, A., and Javitch, J.A. (2001). Symmetrical dimer of the human dopamine transporter revealed by cross-linking Cys-306 at the extracellular end of the sixth transmembrane segment. Proc Natl Acad Sci U S A *98*, 10055-060.

Hastrup, H., Sen, N., and Javitch, J.A. (2003). The human dopamine transporter forms a tetramer in the plasma membrane: cross-linking of a cysteine in the fourth transmembrane segment is sensitive to cocaine analogs. J Biol Chem *278*, 45045-48.

Haussecker, D., and Proudfoot, N.J. (2005). Dicer-dependent turnover of intergenic transcripts from the human beta-globin gene cluster. Mol Cell Biol *25*, 9724-733.

He, J., Mao, C.C., Reyes, A., Sembongi, H., Di Re, M., Granycome, C., Clippingdale, A.B., Fearnley, I.M., Harbour, M., et al. (2007). The AAA+ protein ATAD3 has displacement loop binding properties and is involved in mitochondrial nucleoid organization. J Cell Biol *176*, 141-46.

Heikkila, R.E., Hess, A., and Duvoisin, R.C. (1984). Dopaminergic neurotoxicity of 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine in mice. Science *224*, 1451-53.

Heneka, M.T., Gavrilyuk, V., Landreth, G.E., O'Banion, M.K., Weinberg, G., and Feinstein, D.L. (2003). Noradrenergic depletion increases inflammatory responses in brain: effects on IkappaB and HSP70 expression. J Neurochem *85*, 387-398.

Henry, J.P., Botton, D., Sagne, C., Isambert, M.F., Desnos, C., Blanchard, V., Raisman-Vozari, R., Krejci, E., Massoulie, J., and Gasnier, B. (1994). Biochemistry and molecular biology of the vesicular monoamine transporter from chromaffin granules. J Exp Biol *196*, 251-262.

Hermanson, E., Joseph, B., Castro, D., Lindqvist, E., Aarnisalo, P., Wallén, A., Benoit, G., Hengerer, B., Olson, L., and Perlmann, T. (2003). Nurr1 regulates dopamine synthesis and storage in MN9D dopamine cells. Exp Cell Res *288*, 324-334.

HERTTING, G., and AXELROD, J. (1961). Fate of tritiated noradrenaline at the sympathetic nerve-endings. Nature *192*, 172-73.

Hertzberg, M., Sievertzon, M., Aspeborg, H., Nilsson, P., Sandberg, G., and Lundeberg, J. (2001). cDNA microarray analysis of small plant tissue samples using a cDNA tag target amplification protocol. Plant J *25*, 585-591.

Hieronymus H, Silver PA. (2004). A systems view of mRNP biology. Genes Dev. *18*, 2845-60.

Hillier LD, Lennon G, Becker M, Bonaldo MF, Chiapelli B, Chissoe S, Dietrich N,DuBuque T, Favello A, Gish W, Hawkins M, Hultman M, Kucaba T, Lacy M, Le M, Le N,Mardis E, Moore B, Morris M, Parsons J, Prange C, Rifkin L, Rohlfing T,Schellenberg K, Bento Soares M, Tan F, Thierry-Meg J, Trevaskis E, Underwood K,Wohldman P, Waterston R, Wilson R, Marra M. (1996). Generation and analysis of 280,000 human expressed sequence tags. Genome Res. *6*, 807-28.

Hirotsune, S., Yoshida, N., Chen, A., Garrett, L., Sugiyama, F., Takahashi, S., Yagami, K., Wynshaw-Boris, A., and Yoshiki, A. (2003). An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. Nature *423*, 91-96.

Hirsch E. C., Mouatt A., Faucheux B., Bonnet A. M., Javoy-Agid F., et al. (1992). *Dopamine, tremor, and Parkinson's disease*.Lancet, 340, 125-6

Hirsch, E.C. (1992). Why are nigral catecholaminergic neurons more vulnerable than other cells in Parkinson's disease? Ann Neurol *32 Suppl*, S88-S93.

Hirsch, E.C., and Faucheux, B.A. (1998). Iron metabolism and Parkinson's disease. Mov Disord *13 Suppl 1*, 39-45.

Holland CA, Mayrand S, Pederson T. (1980). Sequence complexity of nuclear and messenger RNA in HeLa cells. J Mol Biol. *138*, 755-78.

Holton, K.L., Loder, M.K., and Melikian, H.E. (2005). Nonclassical, distinct endocytic signals dictate constitutive and PKC-regulated neurotransmitter transporter internalization. Nat Neurosci *8*, 881-88.

Hong, Y.K., Ontiveros, S.D., and Strauss, W.M. (2000). A revision of the human XIST gene organization and structural comparison with mouse Xist. Mamm Genome *11*, 220-24.

Hough, B.R., Smith, M.J., Britten, R.J., and Davidson, E.H. (1975). Sequence complexity of heterogeneous nuclear RNA in sea urchin embryos. Cell *5*, 291-99.

Hu, Z., Cooper, M., Crockett, D.P., and Zhou, R. (2004). Differentiation of the midbrain dopaminergic pathways during mouse development. J Comp Neurol *476*, 301-311.

Huot, P., and Parent, A. (2007). Dopaminergic neurons intrinsic to the striatum. J Neurochem *101*, 1441-47.

Hurst, L.D., and Smith, N.G. (1999). Molecular evolutionary evidence that H19 mRNA is functional. Trends Genet *15*, 134-35.

Hwang, D.Y., Ardayfio, P., Kang, U.J., Semina, E.V., and Kim, K.S. (2003). Selective loss of dopaminergic neurons in the substantia nigra of Pitx3-deficient aphakia mice. Brain Res Mol Brain Res *114*, 123-131.

Hynes, M.A., Poulsen, K., Armanini, M., Berkemeier, L., Phillips, H., and Rosenthal, A. (1994). Neurotrophin-4/5 is a survival factor for

embryonic midbrain dopaminergic neurons in enriched cultures. J Neurosci Res *37*, 144-154.

Iacono, M., Mignone, F., and Pesole, G. (2005). uAUG and uORFs in human and rodent 5'untranslated mRNAs. Gene *349*, 97-105.

Ichimura, T., Isobe, T., Okuyama, T., Takahashi, N., Araki, K., Kuwano, R., and Takahashi, Y. (1988). Molecular cloning of cDNA coding for brain-specific 14-3-3 protein, a protein kinase-dependent activator of tyrosine and tryptophan hydroxylases. Proc Natl Acad Sci U S A *85*, 7084-88.

Ideker T, Galitski T, Hood L. (2001). A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet. *2*, 343-72.

Ikemoto, K., Kitahama, K., Nishimura, A., Jouvet, A., Nishi, K., Arai, R., Jouvet, M., and Nagatsu, I. (1999). Tyrosine hydroxylase and aromatic L-amino acid decarboxylase do not coexist in neurons in the human anterior cingulate cortex. Neurosci Lett *269*, 37-40.

Ikemoto, K., Satoh, K., Kitahama, K., and Maeda, T. (1996). Demonstration of a new dopamine-containing cell group in the primate rostral telencephalon. Neurosci Lett *220*, 69-71.

Imamura T., Yamamoto S., Ohgane J., Hattori N., Tanaka S., Shiota K. (2004). Non-coding RNA directed DNA demethylation of Sphk1 CpG island. Biochem Biophys Res Commun. *322*, 593-600.

Inaba, S., Li, C., Shi, Y.E., Song, D.Q., Jiang, J.D., and Liu, J. (2005). Synuclein gamma inhibits the mitotic checkpoint function and promotes chromosomal instability of breast cancer cells. Breast Cancer Res Treat *94*, 25-35.

Inamdar, N., Arulmozhi, D., Tandon, A., and Bodhankar, S. (2007). Parkinson's disease: genetics and beyond. Curr Neuropharmacol *5*, 99-113.

Ioannidis, J.P. (2008). Calibration of credibility of agnostic genome-wide associations. Am J Med Genet B Neuropsychiatr Genet *147B*, 964-972.

Iversen, L.L. (1971). Role of transmitter uptake mechanisms in synaptic neurotransmission. Br J Pharmacol *41*, 571-591.

Iwaasa, M., Umeda, S., Ohsato, T., Takamatsu, C., Fukuoh, A., Iwasaki, H., Shinagawa, H., Hamasaki, N., and Kang, D. (2002). 1-Methyl-4-phenylpyridinium ion, a toxin that can cause parkinsonism, alters branched structures of DNA. J Neurochem *82*, 30-37.

Iwata, A., Maruyama, M., Kanazawa, I., and Nukina, N. (2001). alpha-Synuclein affects the MAPK pathway and accelerates cell death. J Biol Chem *276*, 45320-29.

Iwata, A., Miura, S., Kanazawa, I., Sawada, M., and Nukina, N. (2001). alpha-Synuclein forms a complex with transcription factor Elk-1. J Neurochem *77*, 239-252.

Iwatsubo, T. (2007). Pathological biochemistry of alpha-synucleinopathy. Neuropathology *27*, 474-78.

Jacobson KB, Arlinghaus HF, Schmitt HW, Sachleben RA, Brown GM, Thonnard N, SloopFV, Foote RS, Larimer FW, Woychik RP, et al. (1991). An approach to the use of stable isotopes for DNA sequencing. Genomics *9*, 51-9.

Jakes, R., Crowther, R.A., Lee, V.M., Trojanowski, J.Q., Iwatsubo, T., and Goedert, M. (1999). Epitope mapping of LB509, a monoclonal antibody directed against human alpha-synuclein. Neurosci Lett *269*, 13-16.

Jakobsen, A.M., Andersson, P., Saglik, G., Andersson, E., Kölby, L., Erickson, J.D., Forssell-Aronsson, E., Wängberg, B., Ahlman, H., and Nilsson, O. (2001). Differential expression of vesicular monoamine transporter (VMAT) 1 and 2 in gastrointestinal endocrine tumours. J Pathol *195*, 463-472.

Janowski, B.A., Hu, J., and Corey, D.R. (2006). Silencing gene expression by targeting chromosomal DNA with antigene peptide nucleic acids and duplex RNAs. Nat Protoc *1*, 436-443.

Jao, C.C., Der-Sarkissian, A., Chen, J., and Langen, R. (2004). Structure of membrane-bound alpha-synuclein studied by site-directed spin labeling. Proc Natl Acad Sci U S A *101*, 8331-36.

Jassen, A.K., Brown, J.M., Panas, H.N., Miller, G.M., Xiao, D., and Madras, B.K. (2005). Variants of the primate vesicular monoamine transporter-2. Brain Res Mol Brain Res *139*, 251-57.

Jeffery, D.R., and Roth, J.A. (1984). Characterization of membrane-bound and soluble catechol-O-methyltransferase from human frontal cortex. J Neurochem *42*, 826-832.

Jellinger, K.A. (1991). Pathology of Parkinson's disease. Changes other than the nigrostriatal pathway. Mol Chem Neuropathol *14*, 153-197.

Jenco, J.M., Rawlingson, A., Daniels, B., and Morris, A.J. (1998). Regulation of phospholipase D2: selective inhibition of mammalian phospholipase D isoenzymes by alpha- and beta- synucleins. Biochemistry *37*, 4901-09.

Jenner, P. (1998). Oxidative mechanisms in nigral cell death in Parkinson's disease. Mov Disord *13 Suppl 1*, 24-34.

Jenner, P. (2003). Oxidative stress in Parkinson's disease. Ann Neurol *53 Suppl 3*, S26-36; discussion S36-8.

Jett JH, Keller RA, Martin JC, Marrone BL, Moyzis RK, Ratliff RL, SeitzingerNK, Shera EB, Stewart CC. (1989). High-speed DNA sequencing: an approach based upon fluorescence detection ofsingle molecules. J Biomol Struct Dyn. 7, 301-9.

Ji, X., Klarmann, G.J., and Preston, B.D. (1996). Effect of human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein on HIV-1 reverse transcriptase activity in vitro. Biochemistry *35*, 132-143.

Jo, E., McLaurin, J., Yip, C.M., St George-Hyslop, P., and Fraser, P.E. (2000). alpha-Synuclein membrane interactions and lipid specificity. J Biol Chem *275*, 34328-334.

Johnson, L.A., Guptaroy, B., Lund, D., Shamban, S., and Gnegy, M.E. (2005). Regulation of amphetamine-stimulated dopamine efflux by protein kinase C beta. J Biol Chem *280*, 10914-19.

Johnson, R.G. (1988). Accumulation of biological amines into chromaffin granules: a model for hormone and neurotransmitter transport. Physiol Rev *68*, 232-307.

Johnson, S.W., and North, R.A. (1992). Two types of neurone in the rat ventral tegmental area and their synaptic inputs. J Physiol *450*, 455-468.

Jones, S., Kornblum, J.L., and Kauer, J.A. (2000). Amphetamine blocks long-term synaptic depression in the ventral tegmental area. J Neurosci *20*, 5575-580.

Jones, S.R., Gainetdinov, R.R., Wightman, R.M., and Caron, M.G. (1998). Mechanisms of amphetamine

action revealed in mice lacking the dopamine transporter. J Neurosci *18*, 1979-986.

Jong, M.T., Gray, T.A., Ji, Y., Glenn, C.C., Saitoh, S., Driscoll, D.J., and Nicholls, R.D. (1999). A novel imprinted gene, encoding a RING zinc-finger protein, and overlapping antisense transcript in the Prader-Willi syndrome critical region. Hum Mol Genet *8*, 783-793.

Kahle, P.J., Neumann, M., Ozmen, L., Muller, V., Jacobsen, H., Spooren, W., Fuss, B., Mallon, B., Macklin, W.B., et al. (2002). Hyperphosphorylation and insolubility of alpha-synuclein in transgenic mouse oligodendrocytes. EMBO Rep *3*, 583-88.

Kaiser P., & Fon E. A. (2007). *Expanding horizons at Big Sky. Symposium on ubiquitin and signaling.*. EMBO Rep. 8, 817-22.

Kalivas, P.W., and Alesdatter, J.E. (1993). Involvement of N-methyl-D-aspartate receptor stimulation in the ventral tegmental area and amygdala in behavioral sensitization to cocaine. J Pharmacol Exp Ther *267*, 486-495.

Kamal, M., Xie, X., and Lander, E.S. (2006). A large family of ancient repeat elements in the human genome is under strong selection. Proc Natl Acad Sci U S A *103*, 2740-45.

Kamatkar, S., Radha, V., Nambirajan, S., Reddy, R.S., and Swarup, G. (1996). Two splice variants of a tyrosine phosphatase differ in substrate specificity, DNA binding, and subcellular location. J Biol Chem *271*, 26755-761.

Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., et al. (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res *14*, 331-342.

Kan, Z., States, D., and Gish, W. (2002). Selecting for functional alternative splices in ESTs. Genome Res *12*, 1837-845.

Kane M.D., Jatkoe T.A., Stumpf C.R., Lu J., Thomas J.D., Madore S.J. (2000). Assessment of the sensitività and specificity of oligonucleotide (50mer) microarrays. Nucleic Acids Res. *28*, 4552-57.

Kanner, B.I., and Schuldiner, S. (1987). Mechanism of transport and storage of neurotransmitters. CRC Crit Rev Biochem *22*, 1-38.

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., and Gingeras, T.R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. Science *296*, 916-19.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484-88.

Kapranov, P., Willingham, A.T., and Gingeras, T.R. (2007). Genome-wide transcription and the implications for genomic organization. Nat Rev Genet *8*, 413-423.

Karbowski, M., Jeong, S.Y., and Youle, R.J. (2004). Endophilin B1 is required for the maintenance of mitochondrial morphology. J Cell Biol *166*, 1027-039.

Karsten, S.L., Van Deerlin, V.M., Sabatti, C., Gill, L.H., and Geschwind, D.H. (2002). An evaluation of tyramide signal amplification and archived fixed and frozen tissue in microarray gene expression analysis. Nucleic Acids Res *30*, E4.

Kaufmann, Y., Milcarek, C., Berissi, H., and Penman, S. (1977). HeLa cell poly(A)- mRNA codes for a subset of poly(A)+ mRNA-directed proteins with an actin as a major product. Proc Natl Acad Sci U S A *74*, 4801-05.

Kawaguchi, Y., and Kubota, Y. (1997). GABAergic cell subtypes and their synaptic connections in rat frontal cortex. Cereb Cortex *7*, 476-486.

Kebabian, J.W., and Calne, D.B. (1979). Multiple receptors for dopamine. Nature *277*, 93-96.

Keys R. A., & Green M. R. (2001). In *Gene expression. The odd coupling.* Nature 413, 583-85

Khan, F.H., Saha, M., and Chakrabarti, S. (2001). Dopamine induced protein damage in mitochondrial-synaptosomal fraction of rat brain. Brain Res *895*, 245-49.

Kholodilov, N.G., Neystat, M., Oo, T.F., Lo, S.E., Larsen, K.E., Sulzer, D., and Burke, R.E. (1999). Increased expression of rat synuclein in the substantia nigra pars compacta identified by mRNA differential display in a model of developmental target injury. J Neurochem *73*, 2586-599.

Kilty, J.E., Lorang, D., and Amara, S.G. (1991). Cloning and expression of a cocaine-sensitive rat dopamine transporter. Science *254*, 578-79.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. (2005). A high-resolution map of active promoters in the human genome. Nature *436*, 876-80.

Kim J, Bhinge AA, Morgan XC, Iyer VR. (2005). Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. Nat Methods 2, 47-53.

Kim, J., Inoue, K., Ishii, J., Vanti, W.B., Voronov, S.V., Murchison, E., Hannon, G., and Abeliovich, A. (2007). A MicroRNA feedback circuit in midbrain dopamine neurons. Science *317*, 1220-24.

Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006). Distinct pathways for snoRNA and mRNA termination. Mol Cell *24*, 723-734.

Kim, T.H., Barrera, L.O., Qu, C., Van Calcar, S., Trinklein, N.D., Cooper, S.J., Luna, R.M., Glass, C.K., Rosenfeld, M.G., et al. (2005). Direct isolation and identification of promoters in the human genome. Genome Res *15*, 830-39.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. (2005). A high-resolution map of active promoters in the human genome. Nature *436*, 876-80.

Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J,Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R,Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, MurakawaK, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi H, Nakai K, Isogai T, Sugano S. (2006). Diversification of transcriptional modulation: large-scale identification andcharacterization of putative alternative promoters of human genes. Genome Res. *16*, 55-65.

Kingsbury, A.E., Daniel, S.E., Sangha, H., Eisen, S., Lees, A.J., and Foster, O.J. (2004). Alteration in alpha-synuclein mRNA expression in Parkinson's disease. Mov Disord *19*, 162-170.

Kirik, D., Rosenblad, C., Burger, C., Lundberg, C.,

247

Johansen, T.E., Muzyczka, N., Mandel, R.J., and Björklund, A. (2002). Parkinson-like neurodegeneration induced by targeted overexpression of alpha-synuclein in the nigrostriatal system. J Neurosci 22, 2780-791.

Kish, S.J., Shannak, K., and Hornykiewicz, O. (1988). Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease. Pathophysiologic and clinical implications. N Engl J Med 318, 876-880.

Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell 109, 145-48.

Kittappa, R., Chang, W.W., Awatramani, R.B., and McKay, R.D. (2007). The foxa2 gene controls the birth and spontaneous degeneration of dopamine neurons in old age. PLoS Biol 5, e325.

Kiyosawa H., Yamanaka I., Osato N., Kondo S., Hayashizaki Y., RIKEN GR Group; GSL Members. (2003). Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. Genome Res. 13, 1324-34.

Klivenyi, P., Siwek, D., Gardian, G., Yang, L., Starkov, A., Cleren, C., Ferrante, R.J., Kowall, N.W., Abeliovich, A., and Beal, M.F. (2006). Mice lacking alpha-synuclein are resistant to mitochondrial toxins. Neurobiol Dis 21, 541-48.

Kordower, J.H., Freeman, T.B., Chen, E.Y., Mufson, E.J., Sanberg, P.R., Hauser, R.A., Snow, B., and Olanow, C.W. (1998). Fetal nigral grafts survive and mediate clinical benefit in a patient with Parkinson's disease. Mov Disord 13, 383-393.

Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G. (2004). Multiple links between transcription and splicing. RNA 10, 1489-98.

Kraytsberg, Y., Kudryavtseva, E., McKee, A.C., Geula, C., Kowall, N.W., and Khrapko, K., (2006). Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. Nat Genet 38, 518-520.

Kress, G.J., and Reynolds, I.J. (2005). Dopaminergic neurotoxins require excitotoxic stimulation in organotypic cultures. Neurobiol Dis 20, 639-645.

Kuhn, D.M., Arthur, R.E., Thomas, D.M., and Elferink, L.A. (1999). Tyrosine hydroxylase is inactivated by catechol-quinones and converted to a redox-cycling quinoprotein: possible relevance to Parkinson's disease. J Neurochem 73, 1309-317.

Kumar, R., Lozano, A.M., Kim, Y.J., Hutchison, W.D., Sime, E., Halket, E., and Lang, A.E. (1998). Double-blind evaluation of subthalamic nucleus deep brain stimulation in advanced Parkinson's disease. Neurology 51, 850-55.

Kumar, R., Lozano, A.M., Montgomery, E., and Lang, A.E. (1998). Pallidotomy and deep brain stimulation of the pallidum and subthalamic nucleus in advanced Parkinson's disease. Mov Disord 13 Suppl 1, 73-82.

Kuusisto, E., Parkkinen, L., and Alafuzoff, I. (2003). Morphogenesis of Lewy bodies: dissimilar incorporation of alpha-synuclein, ubiquitin, and p62. J Neuropathol Exp Neurol 62, 1241-253.

Lacey, M.G., Mercuri, N.B., and North, R.A. (1989). Two cell types in rat substantia nigra zona compacta distinguished by membrane properties and the actions of dopamine and opioids. J Neurosci 9, 1233-241.

Lai, E., Prezioso, V.R., Tao, W.F., Chen, W.S., and Darnell, J.E. (1991). Hepatocyte nuclear factor 3 alpha belongs to a gene family in mammals that is homologous to the Drosophila homeotic gene fork head. Genes Dev 5, 416-427.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Landry, J.R., Mager, D.L., and Wilhelm, B.T. (2003). Complex controls: the role of alternative promoters in mammalian genomes. Trends Genet 19, 640-48.

Lang, A.E. (2007). The progression of Parkinson disease: a hypothesis. Neurology 68, 948-952.

Lang, A.E., and Lozano, A.M. (1998). Parkinson's disease. First of two parts. N Engl J Med 339, 1044-053.

Lang, A.E., and Lozano, A.M. (1998). Parkinson's disease. Second of two parts. N Engl J Med 339, 1130-143.

Larsen, K.E., Schmitz, Y., Troyer, M.D., Mosharov, E., Dietrich, P., Quazi, A.Z., Savalle, M., Nemani, V., Chaudhry, F.A., et al. (2006). Alpha-synuclein overexpression in PC12 and chromaffin cells impairs catecholamine release by interfering with a late step in exocytosis. J Neurosci 26, 11915-922.

Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ, AltschulSF. (2000). SAGEmap: a public gene expression resource. Genome Res. 10, 1051-60.

Lashuel, H.A., Petre, B.M., Wall, J., Simon, M., Nowak, R.J., Walz, T., and Lansbury, P.T. (2002). Alpha-synuclein, especially the Parkinson's disease-associated mutants, forms pore-like annular and tubular protofibrils. J Mol Biol 322, 1089-1102.

Lavoie, B., Smith, Y., and Parent, A. (1989). Dopaminergic innervation of the basal ganglia in the squirrel monkey as revealed by tyrosine hydroxylase immunohistochemistry. J Comp Neurol 289, 36-52.

Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G. (2004). In search of antisense. Trends Biochem Sci 29, 88-94.

Le Moal, M., and Simon, H. (1991). Mesocorticolimbic dopaminergic network: functional and regulatory roles. Physiol Rev 71, 155-234.

Lebel, M., Gauthier, Y., Moreau, A., and Drouin, J. (2001). Pitx3 activates mouse tyrosine hydroxylase promoter via a high-affinity binding site. J Neurochem 77, 558-567.

Lee, F.J., Liu, F., Pristupa, Z.B., and Niznik, H.B. (2001). Direct binding and functional coupling of alpha-synuclein to the dopamine transporters accelerate dopamine-induced apoptosis. FASEB J 15, 916-926.

Lee, H.C., Yin, P.H., Lin, J.C., Wu, C.C., Chen, C.Y., Wu, C.W., Chi, C.W., Tam, T.N., and Wei, Y.H. (2005). Mitochondrial genome instability and mtDNA depletion in human cancers. Ann N Y Acad Sci 1042, 109-122.

Lee, H.J., and Lee, S.J. (2002). Characterization of cytoplasmic alpha-synuclein aggregates. Fibril formation is tightly linked to the inclusion-forming process in cells. J Biol Chem 277, 48976-983.

Lee, H.J., Choi, C., and Lee, S.J. (2002). Membrane-bound alpha-synuclein has a high aggregation propensity and the ability to seed the aggregation of the cytosolic form. J Biol Chem

277, 671-78.

Lee, K.H., Kim, M.Y., Kim, D.H., and Lee, Y.S. (2004). Syntaxin 1A and receptor for activated C kinase interact with the N-terminal region of human dopamine transporter. Neurochem Res *29*, 1405-09.

Leng, Y., Chuang, D.M., (2006). Endogenous alpha-synuclein is induced by valproic acid through histone deacetylase inhibition and participates in neuroprotection against glutamate-induced excitotoxicity. J. Neurosci *26*, 7502-12.

Lehner B., Williams G., Campbell R.D., Sanderson C.M. (2002). Antisense transcripts in the human genome. *18*, 63-5.

Levanon D, Groner Y. (2004). Structure and regulated expression of mammalian RUNX genes. Oncogene *23*, 4211-9.

Lewis, D.A., Melchitzky, D.S., Sesack, S.R., Whitehead, R.E., Auh, S., and Sampson, A. (2001). Dopamine transporter immunoreactivity in monkey cerebral cortex: regional, laminar, and ultrastructural localization. J Comp Neurol *432*, 119-136.

Lewis, D.A., Sesack, S.R., Levey, A.I., and Rosenberg, D.R. (1998). Dopamine axons in primate prefrontal cortex: specificity of distribution, synaptic targets, and development. Adv Pharmacol *42*, 703-06.

Lewy, F.H. (1912). Paralysis agitans. I. Pathologische Anatomie, Handbuch der Neurologie. Bd. III. Verlag Springer, Berlin , 920-933.

Li, J., Adams, L., Schwartz, S.M., and Bumgarner, R.E. (2003). RNA amplification, fidelity and reproducibility of expression profiling. C R Biol *326*, 1021-030.

Li, Q., Peterson, K.R., Fang, X., and Stamatoyannopoulos, G. (2002). Locus control regions. Blood *100*, 3077-086.

Li, W., Lesuisse, C., Xu, Y., Troncoso, J.C., Price, D.L., and Lee, M.K. (2004). Stabilization of alpha-synuclein protein with aging and familial parkinson's disease-linked A53T mutation. J Neurosci *24*, 7400-09.

Liebl, D.J., Morris, C.J., Henkemeyer, M., and Parada, L.F. (2003). mRNA expression of ephrins and Eph receptor tyrosine kinases in the neonatal and adult mouse central nervous system. J Neurosci Res *71*, 7-22.

Lim, K.L., Dawson, V.L., and Dawson, T.M. (2006). Parkin-mediated lysine 63-linked polyubiquitination: a link to protein inclusions formation in Parkinson's and other conformational diseases? Neurobiol Aging *27*, 524-29.

Lindvall, O., and Björklund, A. (1979). Dopaminergic innervation of the globus pallidus by collaterals from the nigrostriatal pathway. Brain Res *172*, 169-173.

Lipshitz, H.D., Peattie, D.A., and Hogness, D.S. (1987). Novel transcripts from the Ultrabithorax domain of the bithorax complex. Genes Dev *1*, 307-322.

Little, K.Y., Elmer, L.W., Zhong, H., Scheys, J.O., and Zhang, L. (2002). Cocaine induction of dopamine transporter trafficking to the plasma membrane. Mol Pharmacol *61*, 436-445.

Liu, C., Fei, E., Jia, N., Wang, H., Tao, R., Iwata, A., Nukina, N., Zhou, J., and Wang, G. (2007). Assembly of lysine 63-linked ubiquitin conjugates by phosphorylated alpha- synuclein implies Lewy body biogenesis. J Biol Chem *282*, 14558-566.

Liu, N., and Baker, H. (1999). Activity-dependent Nurr1 and NGFI-B gene expression in adult mouse olfactory bulb. Neuroreport *10*, 747-751.

Liu QR, Lu L, Zhu XG, Gong JP, Shaham Y, Uhl GR. (2006). Rodent BDNF genes, novel promoters, novel splice variants, and regulation by cocaine. Brain Res *1067*, 1-12.

Liu, Q.R., Walther, D., Drgon, T., Polesskaya, O., Lesnick, T.G., Strain, K.J., de Andrade, M., Bower, J.H., Maraganore, D.M., and Uhl, G.R. (2005). Human brain derived neurotrophic factor (BDNF) genes, splicing patterns, and assessments of associations with substance abuse and Parkinson's Disease. Am J Med Genet B Neuropsychiatr Genet *134B*, 93-103.

Liu, Q.S., Pu, L., and Poo, M.M. (2005). Repeated cocaine exposure in vivo facilitates LTP induction in midbrain dopamine neurons. Nature *437*, 1027-031.

Liu, Y., Peter, D., Roghani, A., Schuldiner, S., Privé, G.G., Eisenberg, D., Brecha, N., and Edwards, R.H. (1992). A cDNA that suppresses MPP+ toxicity encodes a vesicular amine transporter. Cell *70*, 539-551.

Loder, M.K., and Melikian, H.E. (2003). The dopamine transporter constitutively internalizes and recycles in a protein kinase C-regulated manner in stably transfected PC12 cell lines. J Biol Chem *278*, 22168-174.

Lotharius, J., and Brundin, P. (2002). Impaired dopamine storage resulting from alpha-synuclein mutations may contribute to the pathogenesis of Parkinson's disease. Hum Mol Genet *11*, 2395-2407.

Lotharius, J., and Brundin, P. (2002). Pathogenesis of Parkinson's disease: dopamine, vesicles and alpha-synuclein. Nat Rev Neurosci *3*, 932-942.

Lotharius, J., Barg, S., Wiekop, P., Lundberg, C., Raymon, H.K., and Brundin, P. (2002). Effect of mutant alpha-synuclein on dopamine homeostasis in a new human mesencephalic cell line. J Biol Chem *277*, 38884-894.

Lumsden, A., and Krumlauf, R. (1996). Patterning the vertebrate neuraxis. Science *274*, 1109-115.

Lundström, K., Tenhunen, J., Tilgmann, C., Karhunen, T., Panula, P., and Ulmanen, I. (1995). Cloning, expression and structure of catechol-O-methyltransferase. Biochim Biophys Acta *1251*, 1-10.

Luo, L., and O'Leary, D.D. (2005). Axon retraction and degeneration in development and disease. Annu Rev Neurosci *28*, 127-156.

Mahalik, T.J., Finger, T.E., Stromberg, I., and Olson, L. (1985). Substantia nigra transplants into denervated striatum of the rat: ultrastructure of graft and host interconnections. J Comp Neurol *240*, 60-70.

Malenka, R.C., and Bear, M.F. (2004). LTP and LTD: an embarrassment of riches. Neuron *44*, 5-21.

Malenka, R.C., and Nicoll, R.A. (1999). Long-term potentiation--a decade of progress? Science *285*, 1870-74.

Man MZ, Wang X, Wang Y. (2000). POWER_SAGE: comparing statistical tests for SAGE experiments. Bioinformatics *16*, 953-9.

Mancini-Dinardo, D., Steele, S.J., Levorse, J.M., Ingram, R.S., and Tilghman, S.M. (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev *20*, 1268-282.

Mann, D.M. (1983). The locus coeruleus and its possible role in ageing and degenerative disease of the human central nervous system. Mech Ageing Dev *23*, 73-94.

Mann M, Wilm M. (1995). Electrospray mass spectrometry for protein characterization. Trends Biochem Sci. *20*, 219-24.

Marcu KB, Bossone SA, Patel AJ. (1992). Myc function and regulation. Annu Rev Biochem *61*, 809-60.

Margolis, E.B., Hjelmstad, G.O., Bonci, A., and Fields, H.L. (2003). Kappa-opioid agonists directly inhibit midbrain dopaminergic neurons. J Neurosci *23*, 9981-86.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. Nat Rev Neurosci *5*, 793-807.

Maroteaux, L., Campanelli, J.T., and Scheller, R.H. (1988). Synuclein: a neuron-specific protein localized to the nucleus and presynaptic nerve terminal. J Neurosci *8*, 2804-815.

Marra M, Hillier L, Kucaba T, Allen M, Barstead R, Beck C, Blistain A, Bonaldo M,Bowers Y, Bowles L, Cardenas M, Chamberlain A, Chappell J, Clifton S, Favello A, Geisel S, Gibbons M, Harvey N, Hill F, Jackson Y, Kohn S, Lennon G, Mardis E,Martin J, Mila L, McCann R, Morales R, Pape D, Person B, Prange C, Ritter E,Soares M, Schurk R, Shin T, Steptoe M, Swaller T, Theising B, Underwood K, Wylie T, Yount T, Wilson R, Waterston R. (1999). An enciclopedia of mouse genes. Nat Genet. *21*, 191-4.

Marsden, C.D. (1983). Neuromelanin and Parkinson's disease. J Neural Transm Suppl *19*, 121-141.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature *445*, 666-670.

Martinat, C., Bacci, J.J., Leete, T., Kim, J., Vanti, W.B., Newman, A.H., Cha, J.H., Gether, U., Wang, H., and Abeliovich, A. (2006). Cooperative transcription activation by Nurr1 and Pitx3 induces embryonic stem cell maturation to the midbrain dopamine neuron phenotype. Proc Natl Acad Sci U S A *103*, 2874-79.

Martinez-Vicente, M., Talloczy, Z., Kaushik, S., Massey, A.C., Mazzulli, J., Mosharov, E.V., Hodara, R., Fredenburg, R., Wu, D.C., et al. (2008). Dopamine-modified alpha-synuclein blocks chaperone-mediated autophagy. J Clin Invest *118*, 777-788.

Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., et al. (2003). Distribution of NF-kappaB-binding sites across human chromosome 22. Proc Natl Acad Sci U S A *100*, 12247-252.

Masland, R.H. (2004). Neuronal cell types. Curr Biol *14*, R497-R500.

Masson, J., Sagné, C., Hamon, M., and El Mestikawy, S. (1999). Neurotransmitter transporters in the central nervous system. Pharmacol Rev *51*, 439-464.

Matsumura H, Reich S, Ito A, Saitoh H, Kamoun S, Winter P, Kahl G, Reuter M,Kruger DH, Terauchi R. (2003). Gene expression analysis of plant host-pathogen interactions by SuperSAGE. Proc Natl Acad Sci U S A. *100*, 15718-23.

Matsumura H, Reuter M, Krüger DH, Winter P, Kahl G,

Terauchi R. (2008) SuperSAGE. Methods Mol Biol. *387*, 55-70.

Mattick, J.S. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. Bioessays *25*, 930-39.

Mattick, J.S., and Gagen, M.J. (2001). The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. Mol Biol Evol *18*, 1611-630.

Mattick, J.S., and Makunin, I.V. (2005). Small regulatory RNAs in mammals. Hum Mol Genet *14 Spec No 1*, R121- R132.

Mattick, J.S., and Makunin, I.V. (2006). Non-coding RNA. Hum Mol Genet *15 Spec No 1*, R17- R29.

Matzke M.A., Birchler J.A. (2005). RNAi-mediated pathways in the nucleus. Nat Rev Genet. *6*, 24-35.

Maxwell, S.L., Ho, H.Y., Kuehner, E., Zhao, S., and Li, M. (2005). Pitx3 regulates tyrosine hydroxylase expression in the substantia nigra and identifies a subgroup of mesencephalic dopaminergic progenitor neurons during mouse development. Dev Biol *282*, 467-479.

Männistö, P.T., and Kaakkola, S. (1989). New selective COMT inhibitors: useful adjuncts for Parkinson's disease? Trends Pharmacol Sci *10*, 54-56.

McCaffery, P., and Dräger, U.C. (1994). High levels of a retinoic acid-generating dehydrogenase in the meso-telencephalic dopamine system. Proc Natl Acad Sci U S A *91*, 7772-76.

McCLINTOCK, B. (1956). Controlling elements and the gene. Cold Spring Harb Symp Quant Biol *21*, 197-216.

McKeon F. (2004). p63 and the epithelial stem cell: more than status quo? Genes Dev *18*, 465-9.

McLean, P.J., and Hyman, B.T. (2002). An alternatively spliced form of rodent alpha-synuclein forms intracellular inclusions in vitro: role of the carboxy-terminus in alpha- synuclein aggregation. Neurosci Lett *323*, 219-223.

Melikian, H.E. (2004). Neurotransmitter transporter trafficking: endocytosis, recycling, and regulation. Pharmacol Ther *104*, 17-27.

Melikian, H.E., and Buckley, K.M. (1999). Membrane trafficking regulates the activity of the human dopamine transporter. J Neurosci *19*, 7699-7710.

Melino G. (2003). p73, the "assistant" guardian of the genome? Ann N Y Acad Sci *1010*, 9-15.

Meredith, G.E., Farrell, T., Kellaghan, P., Tan, Y., Zahm, D.S., and Totterdell, S. (1999). Immunocytochemical characterization of catecholaminergic neurons in the rat striatum following dopamine-depleting lesions. Eur J Neurosci *11*, 3585-596.

Meyer-Lindenberg, A., Miletich, R.S., Kohn, P.D., Esposito, G., Carson, R.E., Quarantelli, M., Weinberger, D.R., and Berman, K.F. (2002). Reduced prefrontal activity predicts exaggerated striatal dopaminergic function in schizophrenia. Nat Neurosci *5*, 267-271.

Mighell, A.J., Smith, N.R., Robinson, P.A., and Markham, A.F. (2000). Vertebrate pseudogenes. FEBS Lett *468*, 109-114.

Mignone F, Gissi C, Liuni S, Pesole G. (2002). Untranslated regions of mRNAs. Genome Biol *3*, Reviews 0004.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E,

Giannoukos G, Alvarez P,Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, RussC, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-60.

Mikkelsen, T.S., Wakefield, M.J., Aken, B., Amemiya, C.T., Chang, J.L., Duke, S., Garber, M., Gentles, A.J., Goodstadt, L., et al. (2007). Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. Nature *447*, 167-177.

Milcarek, C. (1979). HeLa cell cytoplasmic mRNA contains three classes of sequences: predominantly poly(A)-free, predominantly poly(A)-containing and bimorphic. Eur J Biochem *102*, 467-476.

Milcarek, C., Price, R., and Penman, S. (1974). The metabolism of a poly(A) minus mRNA fraction in HeLa cells. Cell *3*, 1-10.

Miller, G.W., Erickson, J.D., Perez, J.T., Penland, S.N., Mash, D.C., Rye, D.B., and Levey, A.I. (1999). Immunochemical analysis of vesicular monoamine transporter (VMAT2) protein in Parkinson's disease. Exp Neurol *156*, 138-148.

Milner, H.E., Béliveau, R., and Jarvis, S.M. (1994). The in situ size of the dopamine transporter is a tetramer as estimated by radiation inactivation. Biochim Biophys Acta *1190*, 185-87.

Miranda, M., Sorkina, T., Grammatopoulos, T.N., Zawada, W.M., and Sorkin, A. (2004). Multiple molecular determinants in the carboxyl terminus regulate dopamine transporter export from endoplasmic reticulum. J Biol Chem *279*, 30760-770.

Mironov, A.A., Fickett, J.W., and Gelfand, M.S. (1999). Frequent alternative splicing of human genes. Genome Res *9*, 1288-293.

Missale, C., Nash, S.R., Robinson, S.W., Jaber, M., and Caron, M.G. (1998). Dopaminereceptors: from structure to function. Physiol Rev *78*, 189-225.

Miura K. (1981). The cap structure in eukaryotic messenger RNA as a mark of a strand carrying protein information. Adv Biophys. *14*, 205-38.

Modrek, B., and Lee, C. (2002). A genomic view of alternative splicing. Nat Genet *30*, 13-19.

Modrek, B., Resch, A., Grasso, C., and Lee, C. (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Res *29*, 2850-59.

Moldin S. O. (1997). In *The maddening hunt for madness genes.*. Nat Genet. 17, 127-9

Moll, P.R., Duschl, J., and Richter, K. (2004). Optimized RNA amplification using T7-RNA-polymerase based in vitro transcription. Anal Biochem *334*, 164-174.

Moll U.M., Slade N. (2004). p63 and p73: roles in development and tumor formation. Mol Cancer Res *2*, 371-86.

MONTAGU, K.A. (1957). Catechol compounds in rat tissues and in brains of different animals. Nature *180*, 244-45.

Morlando, M., Ballarino, M., Greco, P., Caffarelli, E., Dichtl, B., and Bozzoni, I. (2004). Coupling between snoRNP assembly and 3' processing controls box C/D snoRNA biosynthesis in yeast. EMBO J *23*, 2392-2401.

Morlando, M., Greco, P., Dichtl, B., Fatica, A., Keller, W., and Bozzoni, I. (2002). Functional analysis of yeast snoRNA and snRNA 3'-end formation mediated by uncoupling of cleavage and polyadenylation. Mol Cell Biol *22*, 1379-389.

Morón, J.A., Zakharova, I., Ferrer, J.V., Merrill, G.A., Hope, B., Lafer, E.M., Lin, Z.C., Wang, J.B., Javitch, J.A., et al. (2003). Mitogen-activated protein kinase regulates dopamine transporter surface expression and dopamine transport capacity. J Neurosci *23*, 8480-88.

Mortensen, O.V., and Amara, S.G. (2003). Dynamic regulation of the dopamine transporter.Eur J Pharmacol *479*, 159-170.

Mosharov, E.V., Staal, R.G., Bové, J., Prou, D., Hananiya, A., Markov, D., Poulsen, N., Larsen, K.E., Moore, C.M., et al. (2006). Alpha-synuclein overexpression increases cytosolic catecholamine concentration. J Neurosci *26*, 9304-311.

Mouse Genome Sequencing Consortium, Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Moussa, C.E., Mahmoodian, F., Tomita, Y., and Sidhu, A. (2008). Dopamine differentially induces aggregation of A53T mutant and wild type alpha-synuclein: insights into the protein chemistry of Parkinson's disease. Biochem Biophys Res Commun *365*, 833-39.

Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903-910.

Murphy, D.L., and Wyatt, R.J. (1975). Neurotransmitter-related enzymes in the major psychiatric disorders: I. Catechol-O-methyl transferase, monoamine oxidase in the affective disorders, and factors affecting some behaviorally correlated enzyme activities. Res Publ Assoc Res Nerv Ment Dis *54*, 277-288.

Murray-Zmijewski F, Lane DP, Bourdon JC. (2006). p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress. Cell Death Differ. *13*, 962-72.

Murrell A., Heeson S., Reik W. (2004) Interaction between differentially methylated regione partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat Gnet. *36*, 889-93.

Muthane, U.B., Swamy, H.S., Satishchandra, P., Subhash, M.N., Rao, S., and Subbakrishna, D. (1994). Early onset Parkinson's disease: are juvenile- and young-onset different? Mov Disord *9*, 539-544.

Naderi, A., Ahmed, A.A., Barbosa-Morais, N.L., Aparicio, S., Brenton, J.D., and Caldas, C. (2004). Expression microarray reproducibility is improved by optimising purification steps in RNA amplification and labelling. BMC Genomics *5*, 9.

Nagakubo, D., Taira, T., Kitaura, H., Ikeda, M., Tamai, K., Iguchi-Ariga, S.M., and Ariga, H. (1997). DJ-1, a novel oncogene which transforms mouse NIH3T3 cells in cooperation with ras. Biochem Biophys Res Commun *231*, 509-513.

Narhi, L., Wood, S.J., Steavenson, S., Jiang, Y., Wu, G.M., Anafi, D., Kaufman, S.A., Martin, F., Sitney, K., et al. (1999). Both familial Parkinson's disease mutations accelerate alpha- synuclein

aggregation. J Biol Chem *274*, 9843-46.

Nelson, E.L., Liang, C.L., Sinton, C.M., and German, D.C. (1996). Midbrain dopaminergic neurons in the mouse: computer-assisted mapping. J Comp Neurol *369*, 361-371.

Neumann, M., Kahle, P.J., Giasson, B.I., Ozmen, L., Borroni, E., Spooren, W., Müller, V., Odoy, S., Fujiwara, H., et al. (2002). Misfolded proteinase K-resistant hyperphosphorylated alpha-synuclein in aged transgenic mice with locomotor deterioration and in human alpha-synucleinopathies. J Clin Invest *110*, 1429-439.

Newton, K., Matsumoto, M.L., Wertz, I.E., Kirkpatrick, D.S., Lill, J.R., Tan, J., Dugger, D., Gordon, N., Sidhu, S.S., et al. (2008). Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies. Cell *134*, 668-678.

Newman, M.B., and Bakay, R.A. (2008). Therapeutic potentials of human embryonic stem cells in Parkinson's disease. Neurotherapeutics *5*, 237-251.

Ng P, Wei CL, Sung WK, Chiu KP, Lipovich L, Ang CC, Gupta S, Shahab A, Ridwan A, Wong CH, Liu ET, Ruan Y. (2005). Gene identification signature (GIS) analysis for transcriptome characterizationand genome annotation. Nat Methods 2, 105-11.

Nigg, E.A. (1997). Nucleocytoplasmic transport: signals, mechanisms and regulation. Nature *386*, 779-787.

Nirenberg, M.J., Liu, Y., Peter, D., Edwards, R.H., and Pickel, V.M. (1995). The vesicular monoamine transporter 2 is present in small synaptic vesicles and preferentially localizes to large dense core vesicles in rat solitary tract nuclei. Proc Natl Acad Sci U S A *92*, 8773-77.

Nirenberg, M.J., Vaughan, R.A., Uhl, G.R., Kuhar, M.J., and Pickel, V.M. (1996). The dopamine transporter is localized to dendritic and axonal plasma membranes of nigrostriatal dopaminergic neurons. J Neurosci *16*, 436-447.

Nishie, M., Mori, F., Fujiwara, H., Hasegawa, M., Yoshimoto, M., Iwatsubo, T., Takahashi, H., and Wakabayashi, K. (2004). Accumulation of phosphorylated alpha-synuclein in the brain and peripheral ganglia of patients with multiple system atrophy. Acta Neuropathol *107*, 292-98.

Nishihara, H., Smit, A.F., and Okada, N. (2006). Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res *16*, 864-874.

Njus, D., Kelley, P.M., and Harnadek, G.J. (1986). Bioenergetics of secretory vesicles. Biochim Biophys Acta *853*, 237-265.

Nordström T, Nourizad K, Ronaghi M, Nyrén P. (2000). Method enabling pyrosequencing on double-stranded DNA. Anal Biochem. *282*, 186-93.

Norris, E.H., Giasson, B.I., Hodara, R., Xu, S., Trojanowski, J.Q., Ischiropoulos, H., and Lee, V.M. (2005). Reversible inhibition of alpha-synuclein fibrillization by dopaminochrome-mediated conformational alterations. J Biol Chem *280*, 21212-19.

Nunes, I., Tovmasian, L.T., Silva, R.M., Burke, R.E., and Goff, S.P. (2003). Pitx3 is required for development of substantia nigra dopaminergic neurons. Proc Natl Acad Sci U S A *100*, 4245-250.

Nygaard, V., and Hovig, E. (2006). Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. Nucleic Acids Res *34*, 996-1014.

Nyrén P. (2007). The history of pyrosequencing. Methods Mol Biol. *373*, 1-14.

O'Neill, M.A., Farooqi, I.S., and Wevrick, R. (2005). Evaluation of Prader-Willi Syndrome gene MAGEL2 in severe childhood-onset obesity. Obes Res *13*, 1841-42.

Ohyama, K., Kawano, H., Asou, H., Fukuda, T., Oohira, A., Uyemura, K., and Kawamura, K. (1998). Coordinate expression of L1 and 6B4 proteoglycan/phosphacan is correlated with the migration of mesencephalic dopaminergic neurons in mice. Brain Res Dev Brain Res *107*, 219-226.

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., et al. (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature *420*, 563-573.

Okochi, M., Walter, J., Koyama, A., Nakajo, S., Baba, M., Iwatsubo, T., Meijer, L., Kahle, P.J., and Haass, C. (2000). Constitutive phosphorylation of the Parkinson's disease associated alpha-synuclein. J Biol Chem *275*, 390-97.

Olanow, C.W., Goetz, C.G., Kordower, J.H., Stoessl, A.J., Sossi, V., Brin, M.F., Shannon, K.M., Nauert, G.M., Perl, D.P., et al. (2003). A double-blind controlled trial of bilateral fetal nigral transplantation in Parkinson's disease. Ann Neurol *54*, 403-414.

Oliver B. (2006). *Tiling DNA microarrays for fly genome cartography.*. Nat Genet. 38, 1101-2.

Orlando V, Strutt H, Paro R. (1997). Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods *11*, 205-14.

Orth, M., Tabrizi, S.J., Schapira, A.H., and Cooper, J.M. (2003). Alpha-synuclein expression in HEK293 cells enhances the mitochondrial sensitivity to rotenone. Neurosci Lett *351*, 29-32.

Osato, N., Suzuki, Y., Ikeo, K., and Gojobori, T. (2007). Transcriptional interferences in cis natural antisense transcripts of humans and mice. Genetics *176*, 1299-1306.

Ostrerova, N., Petrucelli, L., Farrer, M., Mehta, N., Choi, P., Hardy, J., and Wolozin, B. (1999). alpha-Synuclein shares physical and functional homology with 14-3-3 proteins. J Neurosci *19*, 5782-791.

Ottobrini L, Ciana P, Biserni A, Lucignani G, Maggi A. (2006). Molecular imaging: a new way to study molecular processes in vivo. Mol Cell Endocrinol. *246*, 69-75.

Padrón, D., Tall, R.D., and Roth, M.G. (2006). Phospholipase D2 is required for efficient endocytic recycling of transferrin receptors. Mol Biol Cell *17*, 598-606.

Pajares MJ, Ezponda T, Catena R, Calvo A, Pio R, Montuenga LM. (2007). Alternative splicing: an emerging topic in molecular and clinical oncology. Lancet Oncol. *8*, 349-57.

Paleologou, K.E., Schmid, A.W., Rospigliosi, C.C., Kim, H.Y., Lamberto, G.R., Fredenburg, R.A., Lansbury, P.T., Fernandez, C.O., Eliezer, D., et al. (2008). Phosphorylation at Ser-129 but not the phosphomimics S129E/D inhibits the fibrillation of alpha-synuclein. J Biol Chem *283*, 16895-6905.

Palfi, S., Leventhal, L., Chu, Y., Ma, S.Y., Emborg, M., Bakay, R., Déglon, N., Hantraye, P., Aebischer, P., and Kordower, J.H. (2002).

Lentivirally delivered glial cell line-derived neurotrophic factor increases the number of striatal dopaminergic neurons in primate models of nigrostriatal degeneration. J Neurosci *22*, 4942-954.

Pandey, N., Schmidt, R.E., and Galvin, J.E. (2006). The alpha-synuclein mutation E46K promotes aggregation in cultured cells. Exp Neurol *197*, 515-520.

Pang, K.C., Frith, M.C., and Mattick, J.S. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genet *22*, 1-5.

Paradis, E., Clement, S., Julien, P., and Ven Murthy, M.R. (2003). Lipoprotein lipase affects the survival and differentiation of neural cells exposed to very low density lipoprotein. J Biol Chem *278*, 9698-9705.

Parent, A., Lavoie, B., Smith, Y., and Bédard, P. (1990). The dopaminergic nigropallidal projection in primates: distinct cellular origin and relative sparing in MPTP-treated monkeys. Adv Neurol *53*, 111-16.

Parent, A., Sato, F., Wu, Y., Gauthier, J., Lévesque, M., and Parent, M. (2000). Organization of the basal ganglia: the importance of axonal collateralization. Trends Neurosci *23*, S20-27.

Parihar, M.S., Parihar, A., Fujita, M., Hashimoto, M., and Ghafourifar, P. (2008). Mitochondrial association of alpha-synuclein causes oxidative stress. Cell Mol Life Sci *65*, 1272-284.

Parish, C.L., and Arenas, E. (2007). Stem-cell-based strategies for the treatment of Parkinson's disease. Neurodegener Dis *4*, 339-347.

Park, S., Püschel, J., Sauter, B.H., Rentsch, M., and Hell, D. (1999). Spatial working memory deficits and clinical symptoms in schizophrenia: a 4-month follow-up study. Biol Psychiatry *46*, 392-400.

Parsons MJ, D'Souza UM, Arranz MJ, Kerwin RW, Makoff AJ. (2004). The -1438A/G polymorphism in the 5-hydroxytryptamine type 2A receptor gene affects promoter activity. Biol Psychiatry *56*, 406-10.

Pasterkamp, R.J., Kolk, S.M., Hellemons, A.J., and Kolodkin, A.L. (2007). Expression patterns of semaphorin7A and plexinC1 during rat neural development suggest roles in axon guidance and neuronal migration. BMC Dev Biol *7*, 98.

Pavel, S. (1993). Dynamics of melanogenesis intermediates. J Invest Dermatol *100*, 162S-65S.

Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci U S A. *91*, 5022-26.

Percheron, G., Yelnik, J., and François, C. (1984). A Golgi analysis of the primate globus pallidus. III. Spatial organization of the striato-pallidal complex. J Comp Neurol *227*, 214-227.

Perez, R.G., Waymire, J.C., Lin, E., Liu, J.J., Guo, F., and Zigmond, M.J. (2002). A role for alpha-synuclein in the regulation of dopamine biosynthesis. J Neurosci *22*, 3090-99.

Perlow, M.J., Freed, W.J., Hoffer, B.J., Seiger, A., Olson, L., and Wyatt, R.J. (1979). Brain grafts reduce motor abnormalities produced by destruction of nigrostriatal dopamine system. Science *204*, 643-47.

Perrin, R.J., Woods, W.S., Clayton, D.F., and George, J.M. (2000). Interaction of human alpha-Synuclein and Parkinson's disease variants with phospholipids. Structural analysis using site-directed mutagenesis. J Biol Chem *275*, 34393-98.

Pesole G, Liuni S, Grillo G, Saccone C. (1998). UTRdb: a specialized database of 5'- and 3'-untranslated regions of eukaryotic mRNAs. Nucleic Acids Res *26*, 192-5.

Peter, D., Jimenez, J., Liu, Y., Kim, J., and Edwards, R.H. (1994). The chromaffin granule and synaptic vesicle amine transporters differ in substrate recognition and sensitivity to inhibitors. J Biol Chem *269*, 7231-37.

Peter, D., Liu, Y., Sternini, C., de Giorgio, R., Brecha, N., and Edwards, R.H. (1995). Differential expression of two vesicular monoamine transporters. J Neurosci *15*, 6179-188.

Petersen, K., Olesen, O.F., and Mikkelsen, J.D. (1999). Developmental expression of alpha-synuclein in rat hippocampus and cerebral cortex. Neuroscience *91*, 651-59.

Phillips, J., and Eberwine, J.H. (1996). Antisense RNA Amplification: A Linear Amplification Method for Analyzing the mRNA Population from Single Living Cells. Methods *10*, 283-88

Phillipson, O.T. (1979). Afferent projections to the ventral tegmental area of Tsai and interfascicular nucleus: a horseradish peroxidase study in the rat. J Comp Neurol *187*, 117-143.

Pifl, C., Giros, B., and Caron, M.G. (1993). Dopamine transporter expression confers cytotoxicity to low doses of the parkinsonism-inducing neurotoxin 1-methyl-4-phenylpyridinium. J Neurosci *13*, 4246-253.

Plant, K.E., Routledge, S.J., and Proudfoot, N.J. (2001). Intergenic transcription in the human beta-globin gene cluster. Mol Cell Biol *21*, 6507-514.

Pleasance ED, Marra MA, Jones SJ. (2003). Assessment of SAGE in transcript identification. Genome Res. *13*, 1203-15.

Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., et al. (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science *276*, 2045-47.

Ponjavic, J., Ponting, C.P., and Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res *17*, 556-565.

Poole, A.C., Thomas, R.E., Andrews, L.A., McBride, H.M., Whitworth, A.J., and Pallanck, L.J. (2008). The PINK1/Parkin pathway regulates mitochondrial morphology. Proc Natl Acad Sci U S A *105*, 1638-643.

Porritt, M.J., Kingsbury, A.E., Hughes, A.J., and Howells, D.W. (2006). Striatal dopaminergic neurons are lost with Parkinson's disease progression. Mov Disord *21*, 2208-211.

Powell J. (2000). SAGE. The serial analysis of gene expression. Methods Mol Biol. *99*, 297-319.

Powner, D.J., Payne, R.M., Pettitt, T.R., Giudici, M.L., Irvine, R.F., and Wakelam, M.J. (2005). Phospholipase D2 stimulates integrin-mediated adhesion via phosphatidylinositol 4-phosphate 5-kinase Igamma b. J Cell Sci *118*, 2975-986.

Pozner A, Lotem J, Xiao C, Goldenberg D, Brenner O, Negreanu V, Levanon D, GronerY. (2007). Developmentally regulated promoter-switch transcriptionally controls Runx1function during

embryonic hematopoiesis. BMC Dev Biol 7, 84.

Prakash, N., and Wurst, W. (2006). Genetic networks controlling the development of midbrain dopaminergic neurons. J Physiol 575, 403-410.

Prakash, N., Brodski, C., Naserke, T., Puelles, E., Gogoi, R., Hall, A., Panhuysen, M., Echevarria, D., Sussel, L., et al. (2006). A Wnt1-regulated genetic network controls the identity and fate of midbrain-dopaminergic progenitors in vivo. Development 133, 89-98.

Prasanth, K.V., and Spector, D.L. (2007). Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes Dev 21, 11-42.

Prensa, L., and Parent, A. (2001). The nigrostriatal pathway in the rat: A single-axon study of the relationship between dorsal and ventral tier nigral neurons and the striosome/ matrix striatal compartments. J Neurosci 21, 7247-260.

Pristupa, Z.B., McConkey, F., Liu, F., Man, H.Y., Lee, F.J., Wang, Y.T., and Niznik, H.B. (1998). Protein kinase-mediated bidirectional trafficking and functional regulation of the human dopamine transporter. Synapse 30, 79-87.

Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, Cocuzza AJ,Jensen MA, Baumeister K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminatingdideoxynucleotides. Science 238, 336-41.

Pronin, A.N., Morris, A.J., Surguchov, A., and Benovic, J.L. (2000). Synucleins are a novel class of substrates for G protein-coupled receptor kinases. J Biol Chem 275, 26515-522.

Pruitt KD, Tatusova T, Maglott DR. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database ofgenomes, transcripts and proteins. Nucleic Acids Res. 33, D501-4.

Puomila K, Simell O, Huoponen K, Mykkänen J. (2007). Two alternative promoters regulate the expression of lysinuric protein intolerance gene SLC7A7. Mol Genet Metab 90, 298-306.

Putzke, J.D., Wharen, R.E., Wszolek, Z.K., Turk, M.F., Strongosky, A.J., and Uitti, R.J. (2003). Thalamic deep brain stimulation for tremor-predominant Parkinson's disease. Parkinsonism Relat Disord 10, 81-88.

Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C., et al. (2006). Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res 16, 11-19.

Reglodi, D., Tamás, A., Lubics, A., Szalontay, L., and Lengvári, I. (2004). Morphological and functional effects of PACAP in 6-hydroxydopamine-induced lesion of the substantia nigra in rats. Regul Pept 123, 85-94.

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J,Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA. (2000). Genome-wide location and function of DNA binding proteins. Science 290, 2306-9.

Reynolds, A.D., Kadiu, I., Garg, S.K., Glanzer, J.G., Nordgren, T., Ciborowski, P., Banerjee, R., and Gendelman, H.E. (2008). Nitrated alpha-synuclein and microglial neuroregulatory activities. J Neuroimmune Pharmacol 3, 59-74.

Rinn JL, Euskirchen G, Bertone P, Martone R,
Luscombe NM, Hartman S, Harrison PM,Nelson FK, Miller P, Gerstein M, Weissman S, Snyder M. (2003). The transcriptional activity of human Chromosome 22. Genes Dev. 17, 529-40.

Ritz, M.C., Lamb, R.J., Goldberg, S.R., and Kuhar, M.J. (1987). Cocaine receptors on dopamine transporters are related to self-administration of cocaine. Science 237, 1219-223.

Rivett, A.J., and Roth, J.A. (1982). Kinetic studies on the O-methylation of dopamine by human brain membrane-bound catechol O-methyltransferase. Biochemistry 21, 1740-42.

Rivett, A.J., Francis, A., and Roth, J.A. (1983). Distinct cellular localization of membrane- bound and soluble forms of catechol-O-methyltransferase in brain. J Neurochem 40, 215-19.

Roberts, R.C., Force, M., and Kung, L. (2002). Dopaminergic synapses in the matrix of the ventrolateral striatum after chronic haloperidol treatment. Synapse 45, 78-85.

Robertson, H.A. (1992). Dopamine receptor interactions: some implications for the treatment of Parkinson's disease. Trends Neurosci 15, 201-06.

Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem. 242, 84-9.

Ronaghi M, Uhlén M, Nyrén P. (1998). A sequencing method based on real-time pyrophosphate. Sciente 281, 363-65.

Rose, G., Gerhardt, G., Strömberg, I., Olson, L., and Hoffer, B. (1985). Monoamine release from dopamine-depleted rat caudate nucleus reinnervated by substantia nigra transplants: an in vivo electrochemical study. Brain Res 341, 92-100.

Rudnick, G., and Wall, S.C. (1992). The platelet plasma membrane serotonin transporter catalyzes exchange between neurotoxic amphetamines and serotonin. Ann N Y Acad Sci 648, 345-47.

Ryu, K.Y., Maehr, R., Gilchrist, C.A., Long, M.A., Bouley, D.M., Mueller, B., Ploegh, H.L., and Kopito, R.R. (2007). The mouse polyubiquitin gene UbC is essential for fetal liver development, cell-cycle progression and stress tolerance. EMBO J 26, 2693-2706.

Runte, M., Hüttenhofer, A., Gross, S., Kiefmann, M., Horsthemke, B., and Buiting, K. (2001). The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. Hum Mol Genet 10, 2687-2700.

Runte, M., Kroisel, P.M., Gillessen-Kaesbach, G., Varon, R., Horn, D., Cohen, M.Y., Wagstaff, J., Horsthemke, B., and Buiting, K. (2004). SNURF-SNRPN and UBE3A transcript levels in patients with Angelman syndrome. Hum Genet 114, 553-561.

Røsok, O., and Sioud, M. (2005). Systematic search for natural antisense transcripts in eukaryotes (review). Int J Mol Med 15, 197-203.

Sabo PJ, Hawrylycz M, Wallace JC, Humbert R, Yu M, Shafer A, Kawamoto J, Hall R, Mack J, Dorschner MO, McArthur M, Stamatoyannopoulos JA. (2004). Discovery of functional noncoding elements by digital analysis of chromatinstructure. Proc Natl Acad Sci U S A. 101, 16837-42.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S,Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, NingZ, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, DavenportR, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J,Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D; International SNP Map Working Group. (2001). A map of human genome sequence variation containing 1.42 million singlenucleotide polymorphisms. Nature *409*, 928-33.

Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., Kinzler, K.W., and Velculescu, V.E. (2002). Using the transcriptome to annotate the genome. Nat Biotechnol *20*, 508-512.

Saito, Y., Kawashima, A., Ruberu, N.N., Fujiwara, H., Koyama, S., Sawabe, M., Arai, T., Nagura, H., Yamanouchi, H., et al. (2003). Accumulation of phosphorylated alpha- synuclein in aging human brain. J Neuropathol Exp Neurol *62*, 644-654.

Salditt-Georgieff, M., Harpold, M.M., Wilson, M.C., and Darnell, J.E. (1981). Large heterogeneous nuclear ribonucleic acid has three times as many 5' caps as polyadenylic acid segments, and most caps do not enter polyribosomes. Mol Cell Biol *1*, 179-187.

Sambrook, J. (1977). Adenovirus amazes at Cold Spring Harbor. Nature *268*, 101-04.

Samii, A., Nutt, J.G., and Ransom, B.R. (2004). Parkinson's disease. Lancet *363*, 1783-793.

Sanger F, Nicklen S, Coulson AR. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. *74*, 5463-7.

SANO, I., GAMO, T., KAKIMOTO, Y., TANIGUCHI, K., TAKESADA, M., and NISHINUMA, K. (1959). Distribution of catechol compounds in human brain. Biochim Biophys Acta *32*, 586-87.

Sasaki, H., and Hogan, B.L. (1994). HNF-3 beta as a regulator of floor plate development. Cell *76*, 103-115.

Sasaki, M., Kaneuchi, M., Sakuragi, N., and Dahiya, R. (2003). Multiple promoters of catechol- O-methyltransferase gene are selectively inactivated by CpG hypermethylation in endometrial cancer. Cancer Res *63*, 3101-06.

Sasaki, M., Shibata, E., Tohyama, K., Takahashi, J., Otsuka, K., Tsuchiya, K., Takahashi, S., Ehara, S., Terayama, Y., and Sakai, A. (2006). Neuromelanin magnetic resonance imaging of locus ceruleus and substantia nigra in Parkinson's disease. Neuroreport *17*, 1215-18.

Sato, Y., Yoshikawa, A., Yamagata, A., Mimura, H., Yamashita, M., Ookata, K., Nureki, O., Iwai, K., Komada, M., and Fukai, S. (2008). Structural basis for specific cleavage of Lys 63- linked polyubiquitin chains. Nature, PMID: 18758443

Satoh, J., and Suzuki, K. (1990). Tyrosine hydroxylase-immunoreactive neurons in the mouse cerebral cortex during the postnatal period. Brain Res Dev Brain Res *53*, 1-5.

Saucedo-Cardenas, O., Quintana-Hau, J.D., Le, W.D., Smidt, M.P., Cox, J.J., De Mayo, F., Burbach, J.P., and Conneely, O.M. (1998). Nurr1 is essential for the induction of the dopaminergic phenotype and the survival of ventral mesencephalic late dopaminergic precursor neurons. Proc Natl Acad Sci U S A *95*, 4013-18.

Saunders, C., Ferrer, J.V., Shi, L., Chen, J., Merrill, G., Lamb, M.E., Leeb-Lundberg, L.M., Carvelli, L., Javitch, J.A., and Galli, A. (2000). Amphetamine-induced loss of human dopamine transporter activity: an internalization-dependent and cocaine-sensitive mechanism. Proc Natl Acad Sci U S A *97*, 6850-55.

Sánchez-Herrero, E., and Akam, M. (1989). Spatially ordered transcription of regulatory DNA in the bithorax complex of Drosophila. Development *107*, 321-29.

Schaefer CB, Ooi SK, Bestor TH, Bourc'his D. (2007). Epigenetic decisions in mammalian germ cells. Sciente *316*, 398-9.

Schein, J.C., Hunter, D.D., and Roffler-Tarlov, S. (1998). Girk2 expression in the ventral midbrain, cerebellum, and olfactory bulb and its relationship to the murine mutation weaver. Dev Biol *204*, 432-450.

Schena M., Shalon D., Davis R.W., Brown P.O. (1995). Quantitative monitorino of gene expression patterns with a complementare DNA microarray. Science *270*, 467-70.

Schnoor M, Voss P, Cullen P, Böking T, Galla HJ, Galinski EA, Lorkowski S. (2004). Characterization of the synthetic compatible solute homoectoine as a potent PCRenhancer. Biochem Biophys Res Commun. *322*, 867-72.

Scott JK, Smith GP. (1990). Searching for peptide ligands with an epitope library. Science 249, 386-90.

Seeman, P. (2002). Atypical antipsychotics: mechanism of action. Can J Psychiatry *47*, 27-38.

Seiden, L.S., Sabol, K.E., and Ricaurte, G.A. (1993). Amphetamine: effects on catecholamine systems and behavior. Annu Rev Pharmacol Toxicol *33*, 639-677.

Selby, G. (1984). The natural history of Parkinson's disease. Aust Fam Physician *13*, 1-3.

Semchuk, K.M., Love, E.J., and Lee, R.G. (1993). Parkinson's disease: a test of the multifactorial etiologic hypothesis. Neurology *43*, 1173-180.

Serpell, L.C., Berriman, J., Jakes, R., Goedert, M., and Crowther, R.A. (2000). Fiber diffraction of synthetic alpha-synuclein filaments shows amyloid-like cross-beta conformation. Proc Natl Acad Sci U S A *97*, 4897-4902.

Sesack, S.R., Hawrylak, V.A., Matus, C., Guido, M.A., and Levey, A.I. (1998). Dopamine axon varicosities in the prelimbic division of the rat prefrontal cortex exhibit sparse immunoreactivity for the dopamine transporter. J Neurosci *18*, 2697-2708.

Sgadò, P., Albéri, L., Gherbassi, D., Galasso, S.L., Ramakers, G.M., Alavian, K.N., Smidt, M.P., Dyck, R.H., and Simon, H.H. (2006). Slow progressive degeneration of nigral dopaminergic neurons in postnatal Engrailed mutant mice. Proc Natl Acad Sci U S A *103*, 15242-47.

Shamoto-Nagai, M., Maruyama, W., Kato, Y., Isobe, K., Tanaka, M., Naoi, M., and Osawa, T. (2003). An inhibitor of mitochondrial complex I, rotenone, inactivates proteasome by oxidative modification and induces aggregation of oxidized proteins in SH-SY5Y cells. J Neurosci Res *74*, 589-597.

Shang, F., Lu, M., Dudek, E., Reddan, J., and Taylor, A. (2003). Vitamin C and vitamin E restore the resistance of GSH-depleted lens cells to H2O2.

Free Radic Biol Med *34*, 521-530.

Sharp, P.A. (1994). Split genes and RNA splicing. Cell *77*, 805-815.

Shastry, B.S. (2001). Parkinson disease: etiology, pathogenesis and future of gene therapy. Neurosci Res *41*, 5-12.

Shimada, S., Kitayama, S., Lin, C.L., Patel, A., Nanthakumar, E., Gregor, P., Kuhar, M., and Uhl, G. (1991). Cloning and expression of a cocaine-sensitive dopamine transporter complementary DNA. Science *254*, 576-78.

Shimura, H., Hattori, N., Kubo, S., Mizuno, Y., Asakawa, S., Minoshima, S., Shimizu, N., Iwai, K., Chiba, T., et al. (2000). Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. Nat Genet *25*, 302-05.

Shults, C.W., and Kimber, T.A. (1992). Mesencephalic dopaminergic cells exhibit increased density of neural cell adhesion molecule and polysialic acid during development. Brain Res Dev Brain Res *65*, 161-172.

Shults, C.W., Hashimoto, R., Brady, R.M., and Gage, F.H. (1990). Dopaminergic cells align along radial glia in the developing mesencephalon of the rat. Neuroscience *38*, 427-436.

Siddiqui, A.S., Khattra, J., Delaney, A.D., Zhao, Y., Astell, C., Asano, J., Babakaiff, R., Barber, S., Beland, J., et al. (2005). A mouse atlas of gene expression: large-scale digital gene-expression profiles from precisely defined developing C57BL/6J mouse tissues and cells. Proc Natl Acad Sci U S A *102*, 18485-490.

Sidhu, A., Wersinger, C., and Vernier, P. (2004). Does alpha-synuclein modulate dopaminergic synaptic content and tone at the synapse? FASEB J *18*, 637-647.

Sillitoe, R.V., and Vogel, M.W. (2008). Desire, disease, and the origins of the dopaminergic system. Schizophr Bull *34*, 212-19.

Simon, H.H., Bhatt, L., Gherbassi, D., Sgadó, P., and Alberí, L. (2003). Midbrain dopaminergic neurons: determination of their developmental fate by transcription factors. Ann N Y Acad Sci *991*, 36-47.

Simon, H.H., Saueressig, H., Wurst, W., Goulding, M.D., and O'Leary, D.D. (2001). Fate of midbrain dopaminergic neurons controlled by the engrailed genes. J Neurosci *21*, 3126-134.

Simons, C., Pheasant, M., Makunin, I.V., and Mattick, J.S. (2006). Transposon-free regions in mammalian genomes. Genome Res *16*, 164-172.

Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., et al. (2003). alpha-Synuclein locus triplication causes Parkinson's disease. Science *302*, 841.

Smidt, M.P., and Burbach, J.P. (2007). How to make a mesodiencephalic dopaminergic neuron. Nat Rev Neurosci *8*, 21-32.

Smidt, M.P., Asbreuk, C.H., Cox, J.J., Chen, H., Johnson, R.L., and Burbach, J.P. (2000). A second independent pathway for development of mesencephalic dopaminergic neurons requires Lmx1b. Nat Neurosci *3*, 337-341.

Smidt, M.P., Smits, S.M., Bouwmeester, H., Hamers, F.P., van der Linden, A.J., Hellemons, A.J., Graw, J., and Burbach, J.P. (2004). Early developmental failure of substantia nigra dopamine neurons in mice lacking the homeodomain gene Pitx3. Development *131*, 1145-155.

Smidt, M.P., van Schaick, H.S., Lanctôt, C., Tremblay, J.J., Cox, J.J., van der Kleij, A.A., Wolterink, G., Drouin, J., and Burbach, J.P. (1997). A homeodomain gene Ptx3 has highly restricted brain expression in mesencephalic dopaminergic neurons. Proc Natl Acad Sci U S A *94*, 13305-310.

Smith, Y., Lavoie, B., Dumas, J., and Parent, A. (1989). Evidence for a distinct nigropallidal dopaminergic projection in the squirrel monkey. Brain Res *482*, 381-86.

Smits, S.M., Ponnio, T., Conneely, O.M., Burbach, J.P., and Smidt, M.P. (2003). Involvement of Nurr1 in specifying the neurotransmitter identity of ventral midbrain dopaminergic neurons. Eur J Neurosci *18*, 1731-38.

Sofic, E., Riederer, P., Heinsen, H., Beckmann, H., Reynolds, G.P., Hebenstreit, G., and Youdim, M.B. (1988). Increased iron (III) and total iron content in post mortem substantia nigra of parkinsonian brain. J Neural Transm *74*, 199-205.

Sorkina, T., Doolen, S., Galperin, E., Zahniser, N.R., and Sorkin, A. (2003). Oligomerization of dopamine transporters visualized in living cells by fluorescence resonance energy transfer microscopy. J Biol Chem *278*, 28274-283.

Spiess, A.N., Mueller, N., and Ivell, R. (2003). Amplified RNA degradation in T7-amplification methods results in biased microarray hybridizations. BMC Genomics *4*, 44.

Spillantini M. G., Schmidt M. L., Lee V. M., Trojanowski J. Q., Jakes R., & Goedert M. (1997). *Alpha-synuclein in Lewy bodies*. Nature. 388, 839-40.

Spillantini, M.G. (1999). Parkinson's disease, dementia with Lewy bodies and multiple system atrophy are alpha-synucleinopathies. Parkinsonism Relat Disord *5*, 157-162.

Spillantini, M.G., Crowther, R.A., Jakes, R., Hasegawa, M., and Goedert, M. (1998). alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. Proc Natl Acad Sci U S A *95*, 6469-473.

Srinivasan, J., and Schmidt, W.J. (2004). Treatment with alpha2-adrenoceptor antagonist, 2-methoxy idazoxan, protects 6-hydroxydopamine-induced Parkinsonian symptoms in rats: neurochemical and behavioral evidence. Behav Brain Res *154*, 353-363.

Stamm, S., Zhu, J., Nakai, K., Stoilov, P., Stoss, O., and Zhang, M.Q. (2000). An alternative- exon database and its statistical analysis. DNA Cell Biol *19*, 739-756.

Stears, R.L., Getts, R.C., and Gullans, S.R. (2000). A novel, sensitive detection system for high-density microarrays using dendrimer technology. Physiol Genomics *3*, 93-99.

Stefanis, L., Larsen, K.E., Rideout, H.J., Sulzer, D., and Greene, L.A. (2001). Expression of A53T mutant but not wild-type alpha-synuclein in PC12 cells induces alterations of the ubiquitin-dependent degradation system, loss of dopamine release, and autophagic cell death. J Neurosci *21*, 9549-560.

Steinmetz, E.J., Conrad, N.K., Brow, D.A., and Corden, J.L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. Nature *413*, 327-331.

Stevens, C.F. (1998). Neuronal diversity: too many cell types for comfort? Curr Biol *8*, R708-710.

Stiewe T, Pützer BM. (2001). p73 in apoptosis. Apoptosis *6*, 447-52.

Stirewalt, D.L., Pogosova-Agadjanyan, E.L., Khalid, N., Hare, D.R., Ladne, P.A., Sala-Torra, O., Zhao, L.P., and Radich, J.P. (2004). Single-stranded linear amplification protocol results in reproducible and reliable microarray data from nanogram amounts of starting RNA. Genomics *83*, 321-331.

Stokes, A.H., Hastings, T.G., and Vrana, K.E. (1999). Cytotoxic and genotoxic potential of dopamine. J Neurosci Res *55*, 659-665.

Stolc, V., Samanta, M.P., Tongprasit, W., Sethi, H., Liang, S., Nelson, D.C., Hegeman, A., Nelson, C., Rancour, D., et al. (2005). Identification of transcribed sequences in Arabidopsis thaliana by using high-resolution genome tiling arrays. Proc Natl Acad Sci U S A *102*, 4453-58.

Strömberg, I., Almqvist, P., Bygdeman, M., Finger, T.E., Gerhardt, G., Granholm, A.C., Mahalik, T.J., Seiger, A., Hoffer, B., and Olson, L. (1988). Intracerebral xenografts of human mesencephalic tissue into athymic rats: immunochemical and in vivo electrochemical studies. Proc Natl Acad Sci U S A *85*, 8331-34.

Strömberg, I., Bygdeman, M., and Almqvist, P. (1992). Target-specific outgrowth from human mesencephalic tissue grafted to cortex or ventricle of immunosuppressed rats. J Comp Neurol *315*, 445-456.

Sugahara Y, Carninci P, Itoh M, Shibata K, Konno H, Endo T, Muramatsu M,Hayashizaki Y. (2001). Comparative evaluation of 5'-end-sequence quality of clones in CAP trapper andother full-length-cDNA libraries. Gene *263*, 93-102.

Sugimoto M, Oohashi T, Ninomiya Y. (1994). The genes COL4A5 and COL4A6, coding for basement membrane collagen chains alpha 5(IV) and alpha 6(IV), are located head-to-head in close proximity on human chromosome Xq22 and COL4A6 is transcribed from two alternative promoters. Proc Natl Acad Sci U S A *91*, 11679-83.

Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., et al. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science *321*, 956-960.

Sulzer, D., and Rayport, S. (1990). Amphetamine and other psychostimulants reduce pH gradients in midbrain dopaminergic neurons and chromaffin granules: a mechanism of action. Neuron *5*, 797-808.

Sulzer, D., Bogulavsky, J., Larsen, K.E., Behr, G., Karatekin, E., Kleinman, M.H., Turro, N., Krantz, D., Edwards, R.H., et al. (2000). Neuromelanin biosynthesis is driven by excess cytosolic catecholamines not accumulated by synaptic vesicles. Proc Natl Acad Sci U S A *97*, 11869-874.

Sulzer, D., Chen, T.K., Lau, Y.Y., Kristensen, H., Rayport, S., and Ewing, A. (1995). Amphetamine redistributes dopamine from synaptic vesicles to the cytosol and promotes reverse transport. J Neurosci *15*, 4102-08.

Sun H, Palaniswamy SK, Pohar TT, Jin VX, Huang TH, Davuluri RV. (2006). MPromDb: an integrated resource for annotation and visualization of mammalian gene promoters and ChIP-chip experimental data. Nucleic Acids Res *34*, D98-103.

Sun M, Hurst LD, Carmichael GG, Chen J. (2005). Evidence for a preferential targeting of 3'-UTRs by cis-encoded natural antisense transcripts. Nucleic Acids Res *33*, 5533-43.

Sun M, Zhou G, Lee S, Chen J, Shi RZ, Wang SM. (2004). SAGE is far more sensitive than EST for detecting low-abundance transcripts. BMC Genomics *5*, 1.

Suzuki Y, Yoshitomo-Nakagawa K, Maruyama K, Suyama A, Sugano S. (1997). Construction and characterization of a full length-enriched and a 5'-end-enrichedcDNA library. Gene *200*, 149-56.

Szumlinski KK, Abernathy KE, Oleson EB, Klugmann M, Lominac KD, He DY, Ron D,During M, Kalivas PW. (2006). Homer isoforms differentially regulate cocaine-induced neuroplasticity. Neuropsychopharmacology *31*, 768-77.

Szymański, M., and Barciszewski, J. (2003). Regulation by RNA. Int Rev Cytol *231*, 197-258.

Taanman, J.W., and Schapira, A.H. (2005). Analysis of the trinucleotide CAG repeat from the DNA polymerase gamma gene (POLG) in patients with Parkinson's disease. Neurosci Lett *376*, 56-59.

Taft, R.J., Pheasant, M., and Mattick, J.S. (2007). The relationship between non-protein- coding DNA and eukaryotic complexity. Bioessays *29*, 288-299.

Takahashi, N., Miner, L.L., Sora, I., Ujike, H., Revay, R.S., Kostic, V., Jackson-Lewis, V., Przedborski, S., and Uhl, G.R. (1997). VMAT2 knockout mice: heterozygotes display reduced amphetamine-conditioned reward, enhanced amphetamine locomotion, and enhanced MPTP toxicity. Proc Natl Acad Sci U S A *94*, 9938-943.

Takei, N., Skoglösa, Y., and Lindholm, D. (1998). Neurotrophic and neuroprotective effects of pituitary adenylate cyclase-activating polypeptide (PACAP) on mesencephalic dopaminergic neurons. J Neurosci Res *54*, 698-706.

Tan JS, Mohandas N, Conboy JG. (2006). High frequency of alternative first exons in erythroid genes suggests a critical role in regulating gene function. Blood *107*, 2557-61.

Tanner, C.M., Ottman, R., Goldman, S.M., Ellenberg, J., Chan, P., Mayeux, R., and Langston, J.W. (1999). Parkinson disease in twins: an etiologic study. JAMA *281*, 341-46.

Tappe A, Klugmann M, Luo C, Hirlinger D, Agarwal N, Benrath J, Ehrengruber MU,During MJ, Kuner R. (2006). Synaptic scaffolding protein Homer1a protects against chronic inflammatory pain. Nat Med *12*, 677-81.

Tappe A., and Kuner R. (2006). Regulation of motor performance and striatal function by synaptic scaffoldingproteins of the Homer1 family. Proc Natl Acad Sci U S A. *103*, 774-9.

Tatton, W.G., and Olanow, C.W. (1999). Apoptosis in neurodegenerative diseases: the role of mitochondria. Biochim Biophys Acta *1410*, 195-213.

Tenhunen, J., Salminen, M., Lundström, K., Kiviluoto, T., Savolainen, R., and Ulmanen, I. (1994). Genomic organization of the human catechol O-methyltransferase gene and its expression from two distinct promoters. Eur J

Biochem *223*, 1049-059.

Thierry-Mieg D., Thierry-Mieg J., (2006). AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biol *7*, Suppl 1: S12-1-14.

Thill G, Castelli V, Pallud S, Salanoubat M, Wincker P, de la Grange P, Auboeuf D, Schächter V, Weissenbach J. (2006). ASEtrap: a biological method for speeding up the exploration of spliceomes. Genome Res. *16*, 776-86.

Thomas, M.J., Malenka, R.C., and Bonci, A. (2000). Modulation of long-term depression by dopamine in the mesolimbic system. J Neurosci *20*, 5581-86.

Thompson, L., Barraud, P., Andersson, E., Kirik, D., and Björklund, A. (2005). Identification of dopaminergic neurons of nigral and ventral tegmental area subtypes in grafts of fetal ventral mesencephalon based on cell morphology, protein expression, and efferent projections. J Neurosci *25*, 6467-477.

Tofaris, G.K., and Spillantini, M.G. (2005). Alpha-synuclein dysfunction in Lewy body diseases. Mov Disord *20 Suppl 12*, S37-S44.

Tofaris, G.K., and Spillantini, M.G. (2007). Physiological and pathological properties of alpha- synuclein. Cell Mol Life Sci *64*, 2194-2201.

Torres, E.M., Rogers, D.C., Annett, L.E., Sirinathsinghji, D.J., and Dunnett, S.B. (1993). A novel population of tyrosine hydroxylase immunoreactive neurones in the basal forebrain of the common marmoset (Callithrix jacchus). Neurosci Lett *150*, 29-32.

Torres, G.E. (2006). The dopamine transporter proteome. J Neurochem *97 Suppl 1*, 3-10.

Torres, G.E., Carneiro, A., Seamans, K., Fiorentini, C., Sweeney, A., Yao, W.D., and Caron, M.G. (2003). Oligomerization and trafficking of the human dopamine transporter. Mutational analysis identifies critical domains important for the functional expression of the transporter. J Biol Chem *278*, 2731-39.

Torres, G.E., Yao, W.D., Mohn, A.R., Quan, H., Kim, K.M., Levey, A.I., Staudinger, J., and Caron, M.G. (2001). Functional interaction between monoamine plasma membrane transporters and the synaptic PDZ domain-containing protein PICK1. Neuron *30*, 121-134.

Toska, K., Kleppe, R., Armstrong, C.G., Morrice, N.A., Cohen, P., and Haavik, J. (2002). Regulation of tyrosine hydroxylase by stress-activated protein kinases. J Neurochem *83*, 775-783.

Troadec, J.D., Marien, M., Darios, F., Hartmann, A., Ruberg, M., Colpaert, F., and Michel, P.P. (2001). Noradrenaline provides long-term protection to dopaminergic neurons by reducing oxidative stress. J Neurochem *79*, 200-210.

Tufarelli C. (2006). The silence RNA keeps: cis mechanisms of RNA mediated epigenetic silencing in mammals. Philos Trans R Soc Lond B Biol Sci. *361*, 67-79.

Tufarelli, C., Frischauf, A.M., Hardison, R., Flint, J., and Higgs, D.R. (2001). Characterization of a widely expressed gene (LUC7-LIKE; LUC7L) defining the centromeric boundary of the human alpha-globin domain. Genomics *71*, 307-314.

Tufarelli, C., Stanley, J.A., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W.G., and Higgs, D.R. (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of

human genetic disease. Nat Genet *34*, 157-165.

Tunbridge, E.M., Lane, T.A., and Harrison, P.J. (2007). Expression of multiple catechol-o-methyltransferase (COMT) mRNA variants in human brain. Am J Med Genet B Neuropsychiatr Genet *144B*, 834-39.

Twelves, D., Perkins, K.S., and Counsell, C. (2003). Systematic review of incidence studies of Parkinson's disease. Mov Disord *18*, 19-31.

Tzivion, G., Luo, Z., and Avruch, J. (1998). A dimeric 14-3-3 protein is an essential cofactor for Raf kinase activity. Nature *394*, 88-92.

Uéda, K., Fukushima, H., Masliah, E., Xia, Y., Iwai, A., Yoshimoto, M., Otero, D.A., Kondo, J., Ihara, Y., and Saitoh, T. (1993). Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. Proc Natl Acad Sci U S A *90*, 11282-86.

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, NarayanV, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D,Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. (2000). A comprehensive analysis of protein-protein interactions in Saccharomycescerevisiae. Nature *403*, 623-7.

Umeda, S., Muta, T., Ohsato, T., Takamatsu, C., Hamasaki, N., and Kang, D. (2000). The D- loop structure of human mtDNA is destabilized directly by 1-methyl-4-phenylpyridinium ion (MPP+), a parkinsonism-causing toxin. Eur J Biochem *267*, 200-06.

Ungerstedt, U. (1971). Striatal dopamine release after amphetamine or nerve degeneration revealed by rotational behaviour. Acta Physiol Scand Suppl *367*, 49-68.

Urbánek, P., Fetka, I., Meisler, M.H., and Busslinger, M. (1997). Cooperation of Pax2 and Pax5 in midbrain and cerebellum development. Proc Natl Acad Sci U S A *94*, 5703-08.

Usdin, T.B., Mezey, E., Chen, C., Brownstein, M.J., and Hoffman, B.J. (1991). Cloning of the cocaine-sensitive bovine dopamine transporter. Proc Natl Acad Sci U S A *88*, 11168-171.

Valente, E.M., Abou-Sleiman, P.M., Caputo, V., Muqit, M.M., Harvey, K., Gispert, S., Ali, Z., Del Turco, D., Bentivoglio, A.R., et al. (2004). Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science *304*, 1158-160.

Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., and Eberwine, J.H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci U S A *87*, 1663-67.

Van Ness, J., Maxwell, I.H., and Hahn, W.E. (1979). Complex population of nonpolyadenylated messenger RNA in mouse brain. Cell *18*, 1341-49.

Varley, J.M., Macgregor, H.C., and Erba, H.P. (1980). Satellite DNA is transcribed on lampbrush chromosomes. Nature *283*, 686-88.

Vartiainen, S., Pehkonen, P., Lakso, M., Nass, R., and Wong, G. (2006). Identification of gene expression changes in transgenic C. elegans overexpressing human alpha-synuclein. Neurobiol Dis *22*, 477-486.

Vaughan, R.A. (2004). Phosphorylation and regulation of psychostimulant-sensitive neurotransmitter transporters. J Pharmacol Exp Ther *310*, 1-7.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW.

(1995). Serial analysis of gene expression. Science *270*, 484-7.

Venkatesh, B., Kirkness, E.F., Loh, Y.H., Halpern, A.L., Lee, A.P., Johnson, J., Dandona, N., Viswanathan, L.D., Tay, A., et al. (2007). Survey sequencing and comparative analysis of the elephant shark (Callorhinchus milii) genome. PLoS Biol *5*, e101.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., et al. (2001). The sequence of the human genome. Science *291*, 1304-351.

Ventura A, Luzi L, Pacini S, Baldari CT, Pelicci PG. (2002). The p66Shc longevity gene is silenced through epigenetic modifications of analternative promoter. J Biol Chem *277*, 22370-6.

Vila, M., Vukosavic, S., Jackson-Lewis, V., Neystat, M., Jakowec, M., and Przedborski, S. (2000). Alpha-synuclein up-regulation in substantia nigra dopaminergic neurons following administration of the parkinsonian toxin MPTP. J Neurochem *74*, 721-29.

Volles, M.J., Lee, S.J., Rochet, J.C., Shtilerman, M.D., Ding, T.T., Kessler, J.C., and Lansbury, P.T. (2001). Vesicle permeabilization by protofibrillar alpha-synuclein: implications for the pathogenesis and treatment of Parkinson's disease. Biochemistry *40*, 7812-19.

VON EULER, U.S., and LISHAJKO, F. (1957). Dopamine in mammalian lung and spleen. Acta Physiol Pharmacol Neerl *6*, 295-303.

Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., et al. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science *320*, 539-543.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N,Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E,Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, RozenS, Hudson TJ, Lipshutz R, Chee M, Lander ES. (1998). Large-scale identification, mapping, and genotyping of single-nucleotidepolymorphisms in the human genome. Science *280*, 1077-82.

Wang, J., Hu, L., Hamilton, S.R., Coombes, K.R., and Zhang, W. (2003). RNA amplification strategies for cDNA microarray experiments. Biotechniques *34*, 394-400.

Wang, L., Duke, L., Zhang, P.S., Arlinghaus, R.B., Symmans, W.F., Sahin, A., Mendez, R., and Dai, J.L. (2003). Alternative splicing disrupts a nuclear localization signal in spleen tyrosine kinase that is required for invasion suppression in breast cancer. Cancer Res *63*, 4724-730.

Wang XJ, Gaasterland T, Chua NH. (2005). Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. Genome Biol *6*, R30.

Wang, Y., Chen, S., Yang, D., and Le, W.D. (2007). Stem cell transplantation: a promising therapy for Parkinson's disease. J Neuroimmune Pharmacol *2*, 243-250.

Wang, Y.M., Gainetdinov, R.R., Fumagalli, F., Xu, F., Jones, S.R., Bock, C.B., Miller, G.W., Wightman, R.M., and Caron, M.G. (1997). Knockout of the vesicular monoamine transporter 2 gene results in neonatal death and supersensitivity to cocaine and amphetamine. Neuron *19*, 1285-296.

Watahiki A, Waki K, Hayatsu N, Shiraki T, Kondo S, Nakamura M, Sasaki D, Arakawa T, Kawai J, Harbers M, Hayashizaki Y, Carninci P. (2004). Libraries enriched for alternatively spliced exons reveal splicing patterns inmelanocytes and melanomas. Nat Methods *1*, 233-9.

Wayment, H.K., Schenk, J.O., and Sorg, B.A. (2001). Characterization of extracellular dopamine clearance in the medial prefrontal cortex: role of monoamine uptake and monoamine oxidase inhibition. J Neurosci *21*, 35-44.

Weber, M.J. (2005). New human and mouse microRNA genes found by homology search. FEBS J *272*, 59-73.

Weihe, E., Depboylu, C., Schütz, B., Schäfer, M.K., and Eiden, L.E. (2006). Three types of tyrosine hydroxylase-positive CNS neurons distinguished by dopa decarboxylase and VMAT2 co-expression. Cell Mol Neurobiol *26*, 659-678.

Weihe, E., Schäfer, M.K., Erickson, J.D., and Eiden, L.E. (1994). Localization of vesicular monoamine transporter isoforms (VMAT1 and VMAT2) to endocrine cells and neurons in rat. J Mol Neurosci *5*, 149-164.

Weinreb, P.H., Zhen, W., Poon, A.W., Conway, K.A., and Lansbury, P.T. (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. Biochemistry *35*, 13709-715.

Weis, K. (1998). Importins and exportins: how to get in and out of the nucleus. Trends Biochem Sci *23*, 185-89.

Wente, S.R. (2000). Gatekeepers of the nucleus. Science *288*, 1374-77.

Werner A., Berdal A. (2005). Natural antisense transcripts: sound or silence? Physiol Genomics *23*, 125-31

Wersinger, C., and Sidhu, A. (2003). Attenuation of dopamine transporter activity by alpha-synuclein. Neurosci Lett *340*, 189-192.

Wersinger, C., Prou, D., Vernier, P., and Sidhu, A. (2003). Modulation of dopamine transporter function by alpha-synuclein is altered by impairment of cell adhesion and by induction of oxidative stress. FASEB J *17*, 2151-53.

Wevrick, R., Kerns, J.A., and Francke, U. (1994). Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. Hum Mol Genet *3*, 1877-882.

Wexler, E.M., and Geschwind, D.H. (2007). Out FOXing Parkinson disease: where development meets neurodegeneration. PLoS Biol *5*, e334.

Whone, A.L., Moore, R.Y., Piccini, P.P., and Brooks, D.J. (2003). Plasticity of the nigropallidal pathway in Parkinson's disease. Ann Neurol *53*, 206-213.

Wictorin, K., Brundin, P., Sauer, H., Lindvall, O., and Björklund, A. (1992). Long distance directed axonal growth from human dopaminergic mesencephalic neuroblasts implanted along the nigrostriatal pathway in 6-hydroxydopamine lesioned adult rats. J Comp Neurol *323*, 475-494.

Wilkinson, K.D., Ventii, K.H., Friedrich, K.L., and Mullally, J.E. (2005). The ubiquitin signal: assembly, recognition and termination. Symposium on ubiquitin and signaling. EMBO Rep *6*, 815-820.

Williams, S.M., and Goldman-Rakic, P.S. (1998). Widespread origin of the primate mesofrontaldopamine system. Cereb Cortex *8*,

321-345.

Willingham, A.T., and Gingeras, T.R. (2006). TUF love for "junk" DNA. Cell *125*, 1215-220.

Willson, C.A., Foster, R.D., Onifer, S.M., Whittemore, S.R., and Miranda, J.D. (2006). EphB3 receptor and ligand expression in the adult rat brain. J Mol Histol *37*, 369-380.

Wilms, H., Rosenstiel, P., Sievers, J., Deuschl, G., Zecca, L., and Lucius, R. (2003). Activation of microglia by human neuromelanin is NF-kappaB dependent and involves p38 mitogen- activated protein kinase: implications for Parkinson's disease. FASEB J *17*, 500-02.

Wise, R.A. (1996). Neurobiology of addiction. Curr Opin Neurobiol *6*, 243-251.

Woolfe, A., and Elgar, G. (2007). Comparative genomics using Fugu reveals insights into regulatory subfunctionalization. Genome Biol *8*, R53.

Wrana, J.L. (1994). H19, a tumour suppressing RNA? Bioessays *16*, 89-90.

Wu G, Nomoto S, Hoque MO, Dracheva T, Osada M, Lee CC, Dong SM, Guo Z, Benoit N, Cohen Y, Rechthand P, Califano J, Moon CS, Ratovitski E, Jen J, Sidransky D,Trink B. (2003). DeltaNp63alpha and TAp63alpha regulate transcription of genes with distinctbiological functions in cancer and development. Cancer Res *63*, 2351-7.

Xiang, C.C., Chen, M., Ma, L., Phan, Q.N., Inman, J.M., Kozhich, O.A., and Brownstein, M.J. (2003). A new strategy to amplify degraded RNA from small tissue samples for microarray studies. Nucleic Acids Res *31*, e53.

Xie, X., Kamal, M., and Lander, E.S. (2006). A family of conserved noncoding elements derived from an ancient transposable element. Proc Natl Acad Sci U S A *103*, 11659-664.

Xin D., Hu L. and Kong X. (2008). Alternative Promoters influence Alternative Splicing at the Genomic Level. Plos ONE *3*, e2377.

Xu, J., Kao, S.Y., Lee, F.J., Song, W., Jin, L.W., and Yankner, B.A. (2002). Dopamine- dependent neurotoxicity of alpha-synuclein: a mechanism for selective neurodegeneration in Parkinson disease. Nat Med *8*, 600-06.

Yang A, Kaghad M, Wang Y, Gillett E, Fleming MD, Dötsch V, Andrews NC, Caput D, McKeon F. (1998). p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. Mol Cell *2*, 305-16.

Yang, Y., Ouyang, Y., Yang, L., Beal, M.F., McQuibban, A., Vogel, H., and Lu, B. (2008). Pink1 regulates mitochondrial dynamics through interaction with the fission/fusion machinery. Proc Natl Acad Sci U S A *105*, 7070-75.

Yasuda, J., and Hayashizaki, Y. (2008). The RNA continent. Adv Cancer Res *99*, 77-112.

Yavich, L., Tanila, H., Vepsäläinen, S., and Jäkälä, P. (2004). Role of alpha-synuclein in presynaptic dopamine recruitment. J Neurosci *24*, 11165-170.

Yelin, R., Steiner-Mordoch, S., Aroeti, B., and Schuldiner, S. (1998). Glycosylation of a vesicular monoamine transporter: a mutation in a conserved proline residue affects the activity, glycosylation, and localization of the transporter. J Neurochem *71*, 2518-527.

Yelnik, J., François, C., Percheron, G., and Heyner, S. (1987). Golgi study of the primate substantia nigra. I. Quantitative morphology and typology of nigral neurons. J Comp Neurol *265*, 455-472.

Yue, Y., Widmer, D.A., Halladay, A.K., Cerretti, D.P., Wagner, G.C., Dreyer, J.L., and Zhou, R. (1999). Specification of distinct dopaminergic neural pathways: roles of the Eph family receptor EphB1 and ligand ephrin-B2. J Neurosci *19*, 2090-2101.

Zahniser, N.R., and Doolen, S. (2001). Chronic and acute regulation of Na+/Cl- -dependent neurotransmitter transporters: drugs, substrates, presynaptic receptors, and signaling systems. Pharmacol Ther *92*, 21-55.

Zamore, P.D., and Haley, B. (2005). Ribo-gnome: the big world of small RNAs. Science *309*, 1519-524.

Zarow, C., Lyness, S.A., Mortimer, J.A., and Chui, H.C. (2003). Neuronal loss is greater in the locus coeruleus than nucleus basalis and substantia nigra in Alzheimer and Parkinson diseases. Arch Neurol *60*, 337-341.

Zarranz, J.J., Alegre, J., Gómez-Esteban, J.C., Lezcano, E., Ros, R., Ampuero, I., Vidal, L., Hoenicka, J., Rodriguez, O., et al. (2004). The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. Ann Neurol *55*, 164-173.

Zavolan, M., Kondo, S., Schonbach, C., Adachi, J., Hume, D.A., Hayashizaki, Y., Gaasterland, T., RIKEN GER Group, and GSL Members (2003). Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. Genome Res *13*, 1290-1300.

Zavolan, M., van Nimwegen, E., and Gaasterland, T. (2002). Splice variation in mouse full- length cDNAs identified by mapping to the mouse genome. Genome Res *12*, 1377-385.

Zecca, L., Fariello, R., Riederer, P., Sulzer, D., Gatti, A., and Tampellini, D. (2002). The absolute concentration of nigral neuromelanin, assayed by a new sensitive method,increases throughout the life and is dramatically decreased in Parkinson's disease. FEBS Lett *510*, 216-220.

Zecca, L., Zucca, F.A., Albertini, A., Rizzio, E., and Fariello, R.G. (2006). A proposed dual role of neuromelanin in the pathogenesis of Parkinson's disease. Neurology *67*, S8-11.

Zetterström, R.H., Solomin, L., Jansson, L., Hoffer, B.J., Olson, L., and Perlmann, T. (1997). Dopamine neuron agenesis in Nurr1-deficient mice. Science *276*, 248-250.

Zhang, Y., Liu, X.S., Liu, Q.R., and Wei, L. (2006). Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. Nucleic Acids Res *34*, 3465-475.

Zhang, Z.X., and Román, G.C. (1993). Worldwide occurrence of Parkinson's disease: an updated review. Neuroepidemiology *12*, 195-208.

Zheng, G., Dwoskin, L.P., and Crooks, P.A. (2006). Vesicular monoamine transporter 2: role as a novel target for drug development. AAPS J *8*, E682-692.

Zhao, H., Hastie, T., Whitfield, M.L., Børresen-Dale, A.L., and Jeffrey, S.S. (2002). Optimization and evaluation of T7 based RNA linear amplification protocols for cDNA microarray analysis. BMC Genomics *3*, 31.

Zhou, Y., Gu, G., Goodlett, D.R., Zhang, T., Pan, C., Montine, T.J., Montine, K.S., Aebersold, R.H., and Zhang, J. (2004). Analysis of alpha-synuclein-associated proteins by quantitative proteomics.

260

J Biol Chem *279*, 39155-164.

Zhou, Z., Luo, M.J., Straesser, K., Katahira, J., Hurt, E., and Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. Nature *407*, 401-05.

Zhu, J., He, F., Wang, J., and Yu, J. (2008). Modeling transcriptome based on transcript- sampling data. PLoS ONE *3*, e1659.

Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques *30*, 892-7.

Zhuo, D., Zhao, W.D., Wright, F.A., Yang, H.Y., Wang, J.P., Sears, R., Baer, T., Kwon, D.H., Gordon, D., et al. (2001). Assembly, annotation, and integration of UNIGENE clusters into the human genome draft. Genome Res *11*, 904-918.

Zimmerman, J.L., Fouts, D.L., and Manning, J.E. (1980). Evidence for a complex class of nonadenylated mRNA in Drosophila. Genetics *95*, 673-691.

Zimprich, A., Müller-Myhsok, B., Farrer, M., Leitner, P., Sharma, M., Hulihan, M., Lockhart, P., Strongosky, A., Kachergus, J., et al. (2004). The PARK8 locus in autosomal dominant parkinsonism: confirmation of linkage and further delineation of the disease-containing interval. Am J Hum Genet *74*, 11-19.

Zwart, R., Sleutels, F., Wutz, A., Schinkel, A.H., and Barlow, D.P. (2001). Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. Genes Dev *15*, 2361-66.

Zweig, R.M., Cardillo, J.E., Cohen, M., Giere, S., and Hedreen, J.C. (1993). The locus ceruleus and dementia in Parkinson's disease. Neurology *43*, 986-991.