Executive summary:

The Chordate central nervous system contains a complex variety of dopaminergic neurons defined according to well-established cytoarchitectural criteria. A specific topographic pattern of degeneration is noted in human post-mortem brains of Parkinson's disease (PD) and in neurochemical animal models suggesting that differentially expressed gene regulatory networks (GRN) may confer susceptibility or protection to dopaminergic cells.

In DOPAMINET we applied a highly interdisciplinary approach to construct complex networks consisting of Transcription Factors, non-coding RNAs and cis-regulatory elements to identify differences and commonalities among subtypes of dopaminergic neurons in three animal models.

This was achieved through the description of gene expression profiles from transgenic mouse and zebrafish, the screening of chordate cis-regulatory elements in zebrafish and ciona, gene network reconstruction and validation of candidate molecules in gene network perturbation experiments.

The definition of the transcriptional landscape of dopaminergic neurons in mouse has provided an unprecedented look at the complexity of the non-coding portion of the transcriptome including repetitive elements and has allowed the discovery of a new functional class of lncRNAs that activates translation. Furthermore, robust evidence for differential expression of Transcription Factors as well as for enrichment of selected Transcription Factor Binding Sites has suggested the existence of different sets of GRNs between mouse dopaminergic subtypes.

In zebrafish we have screened promoters-enhancer combinations for reporter expression in dopaminergic cells. We have also made significant advances towards the understanding of the diversity of dopaminergic neuronal subpopulations by having defined for the first time a complete molecular Transcription Factor code indicative of each major dopaminergic subtype.

In the ascidian Ciona and Fallusia mamillata we have largely increased our knowledge of the repertory of Transcription Factors expressed in dopaminergic neurons and identified some cis-regulatory elements.

By network reconstruction, a list of genes of unknown function has been identified as at the core of the dopaminergic GRN although there was no evidence of a shared mechanism in chordates.

Gene network perturbation experiments in mouse and zebrafish have validated the role of selected Transcription Factors in dopaminergic neurons differentiation. Their use in cell conversion experiments from dermal fibroblasts to dopaminergic cells render them potential molecular tools for manipulating the dopaminergic phenotype in vivo for restorative therapy of PD.

New innovative technologies have been developed and rapidly adopted by the scientific community at large. nanoCAGE allowed the description of the promoters used in mouse and zebrafish dopaminergic cell. A high-content screening pipeline for intelligent automated acquisition of high resolution multi-dimensional data have been used to screen for cis-regulatory elements in zebrafish.

Three start-ups have been generated to fully exploit the translational potential and the impact on society of the innovative basic research carried out in DOPAMINET.

Special efforts have been dedicated to the dissemination of DOPAMINET to scientists and neurologists as well as to patients in collaboration with national PD patients associations. Workshops have been jointly organized with other FP7 European consortia and summer schools have made students aware of the new technologies and approaches in the functional genomics of the nervous system and disease.

Project Context and Objectives:
Parkinson's disease (PD) is a degenerative disorder of the central nervous system (CNS) that is classically defined in terms of motor symptoms consequent to degeneration of mainly A9 dopaminergic neurons in the mesencephalon. No pharmacological treatment is currently available to slow or arrest neurodegeneration. Any treatment that can impact PD has a profound effect on European health since this disease is affecting more than 1.2 million European citizen provoking a large social and economic burden. The aetiology of the disease remains unclear.

Among mesencephalic dopaminergic cell system, A9 neurons are present in the Substantia Nigra (SN) while A10 neurons are mainly confined to the Ventral tegmental Area (VTA). While A9 cells degenerate, in PD A10 neurons are spared. The molecular basis of this selective neurodegeneration is unknown.

A description of commonalities and differences in the gene expression profiles of dopaminergic cell types as well as with non-dopaminergic neurons may provide candidate Gene Regulatory Network (GRNs) for conferring susceptibility or protection in disease. Furthermore, the identification of crucial players in dopaminergic cell differentiation and maintenance may suggest molecular tools for manipulating the dopaminergic phenotype in vivo for restorative therapy of PD.

In DOPAMINET we applied a highly interdisciplinary approach to construct GRNs consisting of Transcription Factors (TF), non-coding RNAs and cis-regulatory elements to identify differences and commonalities among subtypes of dopaminergic neurons in three animal models.

This was achieved through the description of gene expression profiles from transgenic mouse and zebrafish (WP1), the screening of chordate cis-regulatory elements in zebrafish and ciona (WP2), gene network reconstruction (WP3) and validation of candidate molecules in gene network perturbation experiments (WP4).

To reach the objective "Analysis of the Chordate Transcriptome" we needed transgenic lines where dopaminergic neurons were labelled, a technology that allowed the purification of a sufficient number of cells and methods for expression profiling from a small quantity of RNA. Therefore, RIKEN and SISSA developed a new technology, named nanoCAGE, that maps active promoters by identifying the transcription start sites (TSSs) at genome wide-levels. nanoCAGE miniaturizes the requirement for RNA as starting material to the nanogram range and it can be applied to fixed tissues (Plessy et al., Nature methods 2010; Plessy et al., Genome Research 2012). nanoCAGE is a major achievement of DOPAMINET and it is currently used by several laboratories. In DOPAMINET nanoCAGE has been applied to profile mouse and zebrafish dopaminergic subtypes.

Global transcriptome of mouse dopaminergic neurons (SISSA, UCL, RIKEN).

Transgenic mouse GFP-labeled A9 and A10 dopaminergic cells were isolated with LCM. As in D3.1-D3.3, by Affymetrix platform we identified a long list of genes enriched in dopaminergic cells as well as genes differentially expressed between dopaminergic subtypes. TFs enriched in A9 or A10 cells have been listed. These results answered to one of the questions of DOPAMINET proving the existence of different transcriptional networks in anatomically-defined dopaminergic subtypes.

In D1.5-D1.10 nanoCAGE libraries were prepared and more than 18 million tags were extracted. In D3.2-D3.4 we showed that the major part of tags (63%) belonged to coding sequences, 10% of them to non-coding RNA sequences and around 20% to repetitive elements. The remaining tags were not associated to any transcripts and may represent TSSs of genes yet to be characterized. The transcriptional landscape of mouse A9 and A10 neurons have been a major achievement of DOPAMINET representing the first example for an homogeneous population of neurons. In this dataset we have discovered: alternative promoters of alpha synuclein that transcribed mRNAs encoding exclusively for the C-terminal part of the protein (Simone, Carninci, Gustincich, manuscript in preparation); A10 and, in minor part, A9 neurons express a variety of olfactory receptors responding to odorant-like molecules (Zucchelli, Carninci, Gustincich, manuscript in preparation); 439 independently transcribed SINE.ID loci in the genome were found with 21 of them exclusively in the mouse midbrain, some were under the control of the dopaminergic TF Nurr1 (Fedele, Carninci, Gustincich, manuscript in preparation); as in D3.7, 15% of TSS were antisense (AS) to known genes. We thus cloned AS that were associated to genes involved in familial PD such as AS Uchl1, a lncRNA AS to Uchl1/PARK5. AS Uchl1 is the representative member of a new functional class of lncRNAs, named SINEUPs, that are antisense to protein encoding genes and activate their translation (Carrieri et al., Nature 2012). This is the first example of a sequence-specific activator of translation, representing a major achievement of DOPAMINET.

Global transcriptome of zebrafish dopaminergic neurons (ALU-FR, RIKEN).

As shown in D4.3 catecholaminergic neurons were FACS-sorted from transgenic zebrafish strains expressing GFP in several types of catecholaminergic neurons. (Manoli and Driever, 2012, Cold Spring Harbor Protoc.). RNA was either processed for nanoCAGE, or amplified. The full set of Ilumina data has been released to the public by submission to GEO (GSE41373). This is a major contribution to the field since they are the first gene expression profiles of dopaminergic neurons in zebrafish. The major use of these expression profiles in DOPAMINET has been for the identification of TFs expressed in dopaminergic neurons (see below).

Concomitantly, UoB, UCL and RIKEN generated the first vertebrate developmental promoterome using CAGE sequencing and H3k4me3 ChIP sequencing at 12 stages of zebrafish (Nepal et al., submitted; Haberle et al., submitted).

Analysis of gene expression in human (SISSA, RIKEN, UCL).

Although gene expression in human dopaminergic neurons was not among the initial aims of the project, we were conscious that one of the long-term goals of DOPAMINET is the exploitation of results for translational research in PD. Therefore UCL has constructed a database of gene expression profiling in dopaminergic neurons including all published data from PD post-mortem brains. This database has been used in all our GRNs analysis (Taccioli et al, Database 2011). We also took advantage of FANTOM5 dataset from RIKEN. In this project, a simplified CAGE protocol was used to profile over 1000 human and 500 mouse samples to build a promoter-level mammalian expression atlas. In D3.7 we discovered antisense lncRNA transcripts in loci associated to familial PD such as a-synuclein, DJ-1, LRRK2 and MAPT. (Zucchelli, Carninci, Gustincich, manuscript in preparation). Finally, we studied gene expression in the blood of de novo PD patients and controls. Gene expression profiling was

able to discriminate PD patients from age-matched controls. The majority of genes differentially expressed in blood were also present in mouse A9 neurons and human Substantia Nigra.

A crucial objective of DOPAMINET was the "High throughput Screening of cis-regulatory elements"

Here we wanted to identify cis-regulatory elements and transcriptional networks that participate in the coding of dopaminergic neurons identity. To this purpose, we combined different approaches. First we aimed to screen candidate cis-regulatory elements in zebrafish and Ciona embryos, selected by intra-phylum phylogenetic footprinting, taking advantage of fast co-injection-based assays. By gene expression profiling and in situ hybridization experiments we listed combinations of TFs specifically present in selected dopaminergic subtypes. Finally, by identifying the promoterome of dopaminergic cells we searched for TFBSs enrichments. We can thus prioritize TFs to be tested in vivo with Loss-of-Function (LOF) or Gain-of-Function (GOF) experiments.

To analyze dopaminergic neuron-specific activity of putative Core Regulatory Modules (CRMs), KIT and UoB partners established a HTS pipeline for the automated intelligent imaging of zebrafish embryonic brains at high resolution. These tools and technology are being utilized in several additional projects in the field of tissue and whole organism high content screening. The results of the work have been partially published in Peravali, R., Gehrig, J. et al. (2011) BioTechniques 50(5):319-324) and were editorially highlighted in Blow NS. (2011) BioTechniques. 50(5):275. Additionally, a detailed accompanying protocol for intelligent high content screening has been published in the BioTechniques 2012 Protocol Guide. In D2.5, this HTS pipeline has been validated by screening for reporter activity of 202 enhancer-promoter combinations, based on images of thousands of embryos. (Gehirg et al, Nature methods 2009). To this purpose we cloned over 40 zebrafish core promoters for analysis of promoter enhancer interaction specificity. We identified dopaminergic neurons-specific enhancer candidates by phylogenetic footprinting, TFBS analysis and H3k4me1 ChIP sequencing. Over 70 candidate enhancers were functionally tested in transgenic zebrafish.

Transcriptional network in zebrafish (ALU-FR).

In D4.6 ALU-FR developed the Virtual Brain Explorer (ViBE-Z), a software that automatically maps gene expression data with cellular resolution to a 3D standard larval zebrafish (Danio rerio) brain (Ronneberg et al., Nature methods, 2012). This represents a major technological breakthrough. Analyzing gene expression data we have identified 6 new TFs that were validated by whole mount in situ hybridization. These TFs were either coexpressed in A11 neurons or co-expressed with otpa, which is required in precurors of A11-type dopaminergic neurons. We then systematically screened gene expression databases (see http://www.zfin.org online) for TFs expressed in the proximity of ventral diencephalic dopaminergic neurons during zebrafish embryogenesis. A significant number of additional TFs was found and subjected to coexpression analysis. These data suggests that there are combinations of transcription factor families that specify several dopaminergic groups. These include dlx genes, lhx/lim genes, nkx genes, pou3 class genes, and some nuclear orphan receptors in the diencephalon. In the telencephalon it is only etv1. On the other hand, genes like otp, sim1, gata3, otx or zic2a appear to be specific to dopaminergic groups in restricted brain regions only. In summary, we have made significant advances towards the

understanding of the diversity of dopaminergic neuronal subtypes by having defined for the first time a complete molecular transcription factor code indicative of each major dopaminergic subtype in zebrafish.

Transcriptional network in Ciona (CNRS).
The aim to expression profile dopaminergic neurons in Ciona was abandoned early in the project for experimental and theroretical reasons. These made us switched our attention to another ascidian, Phallusia mammillata. We have thus sequenced, assembled and annotated the genome of this animal. While at the beginning of the project only two TFs were known as expressed in dopaminergic neurons, by scanning the Ciona intestinalis expression data section of the ANISEED database we substantially increased the list of Ciona dopaminergic genes to more than 20. The corresponding Phallusia mammillata scaffold included a long 5' flanking sequence for Otx, Meis, SoxB1, Alox12 and Agtr1a/b. For each of these genes, we have identified novel cis-regulatory regions that were conserved in both ascidian species.

Promoterome and cis-regulatory elements in the mouse.
We then identified the promotorome of dopaminergic cells. Promoters were defined as the genomic region around TCs in nanoCAGE dataset. Differentially expressed TCs were identified for A9 and A10 cells as determined after comparison with other dataset and against each other (A9 vs. A10). 1000 TC enriched in A9 neurons were associated to non-CpG while 200 to CpG island. We then looked at the TFBS over-representation of differentially expressed promoters of A9 and of A10 cells. TFBS showing the greatest difference were: EN1, Hand::Tcfe2a, Nurr1, Stat3, Sox5. Furthermore, in an additional analysis, overrepresentation for En1, Nurr1, Pbx, GATA2, NHLH1 and HIF was detected. This was encouraging since En1 and Nurr1 are two crucial TFs involved in A9 neurons differentiation and maintenance.

As defined in the objective "Network Reconstruction" the aim in DOPAMINET was to identify the GRN responsible for the definition of single dopaminergic neurons. In particular, we aimed to assess whether dopaminergic neuron networks are comparable across chordate organisms and what are the differences between subtypes.

To answer these questions YH and UCL generated a database, named ParkDB, that contained all the datasets published on dopaminergic neurons and PD and integrated with DOPAMINET datasets available at the time (Taccioli, DATABASE, 2012). This allowed the identification of key differentially expressed genes. By network reconstruction, we concluded that there is no evidence of a shared mechanism among the different GRNs responsible for the dopaminergic phenotype among the different species. This answers one of the major questions in DOPAMINET. We then focused our attention on the generation of a regulatory network of the mouse. When a more conservative selection of links were adopted, the network presented 83 nodes. While many of the genes known to be associated to Dopamine production and PD were certainly part of the network, the most relevant sub-network, and the group of nodes associated, were unknown genes. The identification of these unknown genes is an important and unexpected contribution of DOPAMINET.

In the objective "Network Perturbations and target identification" we propose to validate by GOF or LOF experiments the role of candidate TFs identified in DOPAMINET. To do so we took advantage of three experimental settings.

Perturbation experiments in zebrafish (ALU-FR).
We have previously defined a set of TFs expressed in selected dopaminergic neurons in zebrafish. Here we performed individual and combinatorial LOF experiments by in vivo knockdown using antisense Morpholino technology. In summary, ALU-FR data established Nkx2 family members as crucial contributors to ventral DA specification in zebrafish (Manoli and Driever, manuscript in preparation). Arx appeared to be essential for proper expression of the DA neurotransmitter phenotype in the prethalamus. Functional inhibition of Isl1 by LOF experiments resulted in a significant decrease of th expression in the prethalamus (Data published in the DOPAMINET manuscript Filippi et al., Developmental biology 2012). This works also established that vDC DA neurons are potentially light-sensing neurons. (Fernades et al. Current Biology 2012). As a major achievement of DOPAMINET, we have thus identified a significant number of epistatic relationships that place crucial TFs into a regulatory hierarchy during ventral diencephalic DA differentiation.

Perturbation experiments in the dopaminergic cell line MN9D (UCL, YH, SISSA, RIKEN).

Since TFBS for Nurr1 was the most enriched motif associated to TSS identified in mouse dopaminergic cells as in D3.4, D3.8, we carried out perturbation experiments by inducible expression of Nurr1 in the dopaminergic-like MN9D cell line. A detailed time course of activation was analyzed vith Illumina microarray and nanoCAGE. To carry out a network analysis we took advantage of ARACNE. Network generated two major clusters: C3 associated to inhibition of neuron-associated genes and C5 associated to mRNA-machinery genes (including chromatin transformation). By using DREM 2.0 the TFs associated to the selected clusters were NKX3-1, PPARA and NFE2L1.

Perturbation experiments in primary dermal fibroblasts (SISSA).

In a collaboration with Dr Vania Broccoli at San Raffaele in Milano, Italy, SISSA aimed to generate dopaminergic neurons through the direct conversion of somatic cells by forced expression of lineage-specific factors including Nurr1. As in Caiazzo et al., Nature 2011, Nurr1, in combination with Mash1 and Lmx1a, were able to generate directly functional dopaminergic neurons from mouse and human fibroblasts without reverting to a progenitor cell stage. Microarrays analysis carried out at SISSA showed that these cells clustered with A9 and A10 adult mesencephalic dopaminergic (mDA) neurons (as shown in D1.9). We then took advantage of the cell conversion assay to test the activity of TFs found expressed in dopaminergic neurons and which TFBS was enriched around TSS obtained with nanoCAGE as shown in D3.8. Importantly, lentiviral delivery of Etv5, HIF and PBX3 strongly increased the number of TH-positive cells. We are currently gene expression profiling these cells to monitor whether they resemble a specific mesencephalic dopaminergic subtype type.

These results demonstrate the validity of the DOPAMINET approach to identify TFs able to trigger dopaminergic cells' differentiation.

IP transfer
Three start-ups have been generated to fully exploit the translational potential and the impact on society of the innovative basic research carried out in DOPAMINET:

TranSINE Technology is dedicated to use SINEUPs to increase protein levels in protein manufactoring and RNA therapeutics for haploinsufficiences (see http://www.transsine.com online).
During the development of the image based screening technology for DOPAMINET a strong market potential was identified. As a consequence, the spin-off company "Acquifer" was founded in the beginning of 2012 (see http://www.acquifer.de online).
By the identification of a gene signature for PD diagnosis from blood, the start-up PARKscreen has been founded.

Dissemination
Special efforts have been dedicated to the dissemination of DOPAMINET to scientists and neurologists as well as to patients in collaboration with national PD patients associations. Workshops have been jointly organized with other Seventh Framework Programme (FP7) European consortia and summer schools have made students aware of the new technologies and approaches in the functional genomics of the nervous system and disease.

Project Results:
The central concept of DOPAMINET starts from the observation that the Central Nervous System (CNS) contains an enormous variety of cell types that give rise to homogeneous neuronal cell groups organized into intricate and complex networks. The lack of adequate cell type-specific markers renders the task of isolating these cells very challenging. This heterogeneity has a profound impact in the description of cellular transcriptome since RNAs expressed only in selected populations become diluted out and undetectable when studied from total brain tissue preparation. Furthermore, this complexity renders very difficult to have a complete lists of all the cell types of a specific area of the brain or even of a single neuronal network. The ability to identify all the types of neurons and describe their transcriptional profiles is important for our understanding of the functional organization of the nervous system according to a "bottom up" approach. This entails a description of all their chemical mediators, channels and receptors through functional genomics techniques. Furthermore, the identification of their repertory of transcription factors (TFs) and other regulatory modules may lead to an understanding of the gene networks involved in their differentiation and maintenance.

This is also important for neurodegenerative diseases where a major open question concerns the selective susceptibility of specific neuronal cell types.

Parkinson's disease (PD) is a slowly progressive degenerative disorder of the central nervous system (CNS) that is classically defined in terms of motor symptoms consequent to degeneration of specific subsets of mesencephalic dopaminergic cells within Substantia Nigra (SN). Although dopaminergic drugs are effective in alleviating motor symptoms in PD patients, no pharmacological treatment is currently available to slow or arrest the neurodegenerative process. In mammals the mesencephalic dopaminergic cell system contains at least 13 different nuclei as classified in terms of projecting neurons according to well-established cytoarchitectural criteria. Among them, the nigral neurons (A9) are present in the SN while A10 DA neurons are mainly confined to the Ventral tegmental Area (VTA). These nuclei give rise to the mesocortical and mesostriatal projections and are well known for their roles in initiation of movement, reward behaviour and neurobiology of addiction. In post mortem brains of PD patients, a specific topographic pattern of

degeneration is noted with maximal cell loss in A9 neurons and negligible in the A10 medial group.

A description of commonalities and differences between gene expression profiles of dopaminergic cell types as well as with non-dopaminergic neurons may provide candidate molecular pathways for conferring susceptibility or protection in disease. Furthermore, the identification of crucial players in dopaminergic cell differentiation and maintenance may suggest molecular tools for manipulating the dopaminergic phenotype in vivo for restorative therapy of PD.

To be able to fully describe the transcriptional landscape of even a single neuronal cell type, a series of technological advancements are required. First, we need to specifically label these neurons. Then a technology is needed to purify to homogeneity this type of cells. However, although reaching this goal is almost impossible for the majority of neuronal cell types, the most difficult technological challenges are still ahead. In particular, there is the need to purify intact RNAs from these cells and to apply a technology to describe their transcriptome in a reproducible and sensitive manner.

Here it is important to consider the staggering speed of technological advancement. When DOPAMINET was conceived, the large majority of gene expression profiling data was produced taking advantage of the microarray platform. In the last four years next generation sequencing has become of age due to the impressive decrease of the cost of sequencing. This has also heavily impacted our understanding of the mammalian transcriptome. First, the large majority of the genome is transcribed due to the regulated expression of repetitive elements. Then, current estimates suggest that there might be at least as many long non-coding RNAs as protein-encoding genes. Interestingly, many genes present Antisense (AS) transcription on the other filament suggesting the existence of S/AS pairs in a wide range of organisms from bacteria to human. The current knowledge of the role of repetitive elements and lncRNAs in brain function is negligible.

Meantime, system biology and bioinformatics have shown remarkable progress allowing the extrapolation of knowledge from complex and diverse dataset. This has led to the ability of modeling Gene Regulatory Networks (GRNs). These form the core of the developmental programme and/or of the cellular response to environmental signalling. The nodes of these networks are the genes for transcription factors and signaling ligands. Their edges are the regulatory relationships that link these genes. GRNs can be reconstructed by the integration of gene expression profiles, the effect of alterations of gene function, and cis-regulatory module analysis.

It has been postulated that GRNs may be evolutionary conserved. Therefore the use of different but complementary model systems may strongly increase our understanding of GRNs. Model organisms have become increasingly powerful tools to explore fundamental questions in biology, since all most common model organisms have fully sequenced genomes as well as a variety of other high-throughput functional genomics data, enabling a deeper understanding of their biology. Each model organism offers specific experimental advantages, and thus they can be regarded as complementary tools in the quest for biological knowledge of an endogenous process or a disease. For example, together with mouse, zebrafish has become a widely utilized model, owing to several reasons,

among which the ease of morpholino-based Loss-of-Function (LOF) assays, its transparent embryos and fully sequenced genome. Zebrafish has also become the organism of choice for testing cis-regulatory elements in medium throughput by co-injection assays. The ascidian Ciona intestinalis has proven remarkably powerful to reconstruct regulatory networks owing to its deep genomics resources, efficient loss/gain of function assays and remarkably simple body plan.

As a representative example of a functional genomics approach to neural network, DOPAMINET aims to identify crucial GRNs of dopaminergic neurons in Chordates. This is achieved through the following steps:
1. Gene expression profiling of dopaminergic neurons from transgenic mouse, zebrafish and Ciona (WP1).
2. Screening of chordate cis-regulatory elements in zebrafish (WP2).
3. Data analysis and network reconstruction (WP3).
4. Network perturbations (WP4).

This interdisciplinary approach has been implemented thanks to the development of new technologies and their immediate applications to the study of dopaminergic neurons.

As a major objective of the project, the first step was the assembly of a robust gene expression dataset for dopaminergic neurons. To fulfill this aim, DOPAMINET had to develop a series of tools and technologies. Dopaminergic neurons from different species must be labeled and purified near homogeneity. Then, a sensitive and reproducible method needed to be optimized for gene expression profiling from small quantity of fixed, starting material.

To our knowledge, a complete description of the transcriptional landscape of an homogeous neuronal cell population has never been achieved so far, therefore this represented a major challenge and potential advancement of knowledge. In particular, we believed that a meaningful dataset must not be limited to oligonucleotide-based, protein coding enriched microarrays but it must fully exploit the power of next generation sequencing. Special emphasis had to focus on the identification of the repertory of TFs as well as of Transcription Start Sites (TSSs) and associated promoters. An unprecedented look at the non-coding portion of the transcriptome may provide new regulatory elements for cell differentiation and maintenence.

To reach the objective "Analysis of the Chordate Transcriptome" we had to collect dopaminergic neurons from different organisms and carry out gene expression profiling both with Affymetrix microarray and second generation sequencing.

To this purpose we needed transgenic lines where dopaminergic neurons were labelled, a technology that allowed the purification of a sufficient number of cells and methods for expression profiling from a small quantity of RNA. Different strategies have been used to label dopaminergic neurons in different Chordates and for cells' purification. Importantly, we first developed a new technology, named nanoCAGE, to profiles purified cells.

Development of new technologies: nanoCAGE (RIKEN, SISSA).
The Cap Analysis Gene Expression (CAGE) technology was previously developed for the systematic analysis of TTSs in eukaryotic cells and tissues. CAGE is based on sequencing the 5'ends of mRNAs, of which the

integrity is inferred by the presence of their cap. The sequences—which we refer here to as tags—are sufficiently long to be aligned in most cases at a single position of the genome. The first position of this alignment identifies a base pair where transcription is initiated defining a TSS. The number of times a given tag is represented in a library gives an estimate of the expression level of the corresponding transcript. CAGE has previously enabled to map TFBSs in promoters, and to identify LINEs as a source of alternative promoters for protein coding genes. Interestingly, by a large-scale use of CAGE technology as in the FANTOM projects, two main types of promoters were identified. Single dominant peak class promoters (SP) were giving a single dominant TSS and were generally associated with TATA-boxes. General broad distribution (BR) promoters had broad distribution of TSSs generally spread over 100 nt and were strongly associated with CpG islands and were GC rich.

To expand this approach to tiny amounts of ex vivo tissue and to the polyA- fraction of RNAs, RIKEN and SISSA developed nanoCAGE, a technology that miniaturizes the requirement of CAGE for RNA material to the nanogram range and which can be used on fixed tissues (Plessy et al., Nature methods 2010). As proof of principle, nanoCAGE has been applied by RIKEN and SISSA to identify the entire repertory of TSSs in the mouse olfactory epithelium (MOE). Whole mouse MOE was purified with laser capture microdissection (LCM) from fixed histological sections. nanoCAGE analysis revealed the map and architecture of promoters for 87.5% of the mouse olfactory receptors genes, as well as the expression of many novel noncoding RNAs including antisense transcripts (Plessy et al., Genome Research 2012).

nanoCAGE is a major achievement of DOPAMINET and it is currently used by several laboratories around the world. The technology has been patented.

Global transcriptome of mouse dopaminergic neurons (SISSA, RIKEN, UCL).

In D1.1-D1.6 A9 and A10 dopaminergic cells were isolated by taking advantage of Th-GFP 21-22 transgenic mice that selectively express the green fluorescent protein (GFP) in catecholaminergic cells under the control of tyrosine hydroxylase (TH) promoter. In this mouse line the majority of mDA neurons were identified for their eGFP labeling. Furthermore, A9 neurons were distinguished from A10 for their anatomical localization. After fixation with Zinc-fix, a procedure that maintained both tissue and RNA integrity, LCM and pressure catapulting were performed to separately harvest the two populations of neurons near homogeneity. Six independent harvests of 2000 cells each for each type of neuron were completed.

First we profiled these cells by taking advantage of the Affymetrix microarray platform. In D1.4-D1.9 experiments were successfully carried out at SISSA and the bioinformatics analysis at UCL. This aimed to identify genes that were differentially expressed between A9, A10 and mesencephalon in vivo. In D3.1-D3.3 the mesencephalon vs A9/10 comparison resulted in a high number of differentially expressed genes (1285 genes with an adjusted p-value less than 0.01 without fold change filter). Reassuringly TH, the rate limiting enzyme for dopamine biosynthesis and the marker for these cells, appears the very top gene in the list. Similarly, other known key genes involved in dopaminergic neuron differentiation and function were also found amongst top genes, such as Nurr1 (Nr4a2), RET as well as the dopamine plasmalemma and vescicular transporters (DAT, shown as SLC6A3 and VMAT2 as SLC18A2). Importantly, as

shown in D3.8, a list of TFs enriched in DA cells has been compiled. These results are important for neuroscientists at large since we identified as expressed 35 channels, 106 protein kinases, 33 phosphatases, 97 genes for secreted proteins and 56 receptors.

The analysis of A9 vs A10 neurons yielded a smaller list of genes that were significantly differentially expressed at appreciable fold change. Among them, we identified a series of TFs, as LMO4 for A9, that were highly enriched specifically in a dopaminergic cell group. These results answered to one of the major questions of DOPAMINET proving the existence of different transcriptional networks in anatomically defined dopaminergic subtypes.

Interestingly, genomic distribution of the differentially expressed genes showed a significant enrichment for chr2q37 (p-value 2.28 e-8) and chr2q35 (p-value 3.82 e-6) regions. The potential relevance of these findings would require further investigation based on epigenomics analyses (such as CNV, DNA Methylation and histone mark analysis).

In D1.5-D1.10 nanoCAGE libraries were prepared using two biological replicates of mouse A9 and A10 dopaminergic neurons. They were sequenced and more than 18 million tags were extracted. To our knowledge, this is the first attempt to describe the transcriptional landscape of a specific type of neurons.

In D3.2-D3.4 to characterize the transcriptome of dopaminergic neurons in details, we associated mapped tags to RefSeq transcripts and FANTOM3 full length non-coding RNAs. The major part of tags (63%) belonged to coding sequences, 10% of them to non-coding RNA sequences and around 20% to repetitive elements (Multiple Mapping unrescued tags). The remaining tags were not associated to any transcripts and may represent TSSs of genes yet to be characterized. Distribution of tags over the proximal promoter as well as intron/exon structures, 3'ends and intergenic distribution was determined. Tags from the entire library were then aggregated into tag clusters (TCs) when they mapped on the same genome strand and they were at most 27 nucleotides apart. A TPM (tag per million) score was associated to each TC as a direct count of the frequencies of a given clustered group of tags for each million of read sequences. The TPM thus represents an assessment of the expression level of a given transcript starting at a specific TSS. (Lazarevic, Carninci, Gustincich, manuscript in preparation).

The transcriptional landscape of mouse A9 and A10 neurons have been a major achievement of DOPAMINET representing the first example for an homogeneous population of neurons.

In this analysis we made two original observations that may have an impact on PD.

1. A9 and A10 cells present two independent TCs at the 'lpha-synuclein locus (mScna) indicating the presence of at least two major TSSs used differently in the two cell types. Furthermore, an unexpected TC is present within the canonical 4th exon. 5' and 3' RACE coupled to 454 sequencing confirmed nanoCAGE data and cloned previously unnoticed splicing forms without the ATG-containing 2nd exon. When these splicing forms and the 5' truncated transcript starting at exon 4 were transiently expressed, they led to the production of a 42 aa long C-terminal portion of alpha-synuclein. While C-terminal synuclein peptides are commonly

found in post-mortem brains, they are believed to be produced by caspase-mediated protein cleavage. Here we made the original observation that there are two additional mechanisms to synthesize them: a spliced mRNA without the second exon starting from the canonical distal promoter and an mRNA transcribed from a TSS in exon 4. This is important for PD pathogenesis since the C-terminal is crucial in 'lpha-synuclein protein stabilization and fibrillation. Expression of these short peptides may thus interfere with its aggregation and be involved in PD pathogenesis. (Simone, Carninci, Gustincich, manuscript in preparation).
2. NanoCAGE analysis revealed TSSs in 19 loci for olfactory receptor genes. The majority of them were expressed exclusively in A10 neurons initiating in very similar, if not identical, genomic position as in the olfactory epithelium. Expression for some of them was validated by RT-PCR, in situ hybridization and western. Furthermore, after cloning their full-length cDNAs from midbrain, we determined their ligand specificity in response to odors. For three of them carvone was acting as an odor-like agonist. Primary mDA neurons showed Ca2+ responsiveness upon odor mix and carvone stimulation. Homologous human ORs have been cloned from human mesencephalon and proved down-regulated in PD post-mortem brains. (Zucchelli, Carninci, Gustincich, manuscript in preparation).

Non-coding RNA transcriptome: Repetitive elements.

We then focused our attention on the non-coding portion of the transcriptome. Systematic 5'end analysis showed 18% of tags in A9 DA cells' libraries were retrotransposons. This amount was similar in A10 neurons (19%). Members of the LTR family, such as LTR.ERVK, LTR.ERVL and MaLR, represented 5% of tags while LINE sequences 3% of tags. Interestingly, SINE sequences were the largest portion of repetitive elements (7% of the total transcriptome). BC1 is the majorly expressed SINE.ID transcript in the mouse brain. We thus hypothesized that tags mapping to SINE.ID elements were associated to this transcriptional unit. Surprisingly, alignment of SINE.ID tags on the mouse genome indicated that the large majority of these elements is mapping on different genomic positions. Thanks to an extensive 3'RACE analysis, we showed that BC1 transcript represented only 20% of mapped sequences while there were 439 independently transcribed SINE.ID loci in the genome. Among them, 21 were detected as expressed exclusively in the mouse midbrain. Interestingly, they were part of the transcriptional network of Nurr1, a major regulator of dopaminergic neurons differentiation and maintenance since they were induced when Nurr1 was overexpressed in the dopaminergic MN9D cell line. Expressed SINE.IDs showed conserved internal A and B boxes for RNPIII binding and a TATA box at -12. Importantly, the flanking genomic regions were able to drive neuron-specific transcription. (Fedele, Carninci, Gustincich, manuscript in preparation).

Non-coding RNA transcriptome: Antisense Transcription.

As in D3.7, analysis of nanoCAGE libraries of A9 and A10 nuclei also showed that 15% of TSS are antisense (AS) to known genes. We thus cloned and analysed in details AS that were associated to genes involved in familial PD. We identified AS Uchl1 as a nuclear-enriched lncRNA AS to Uchl1/PARK5. AS Uchl1 is selectively expressed in A9 and A10 neurons in vivo. Its transcription is induced by Nurr1 over-expression in the dopaminergic MN9D cell line through selective binding of Nurr1 to AS Uchl1 promoter. AS Uchl1 increases UchL1 protein synthesis at post-transcriptional level, identifying a new functional class of lncRNAs. AS Uchl1 activity depends on the presence of a 5' overlapping sequence and

an embedded inverted SINEB2 element. In addition, mTORc1 inhibition by rapamycin causes an induction of UchL1 protein that is concomitant to shuttling of AS Uchl1 RNA from the nucleus to the cytoplasm and an increased association of the overlapping sense protein-encoding mRNA to active polysomes for translation. Thus, AS Uchl1 is the first lncRNA able to stimulate translation of specific mRNAs, in conditions in which CAP-dependent translation is reduced.

We thus proposed a model where these lncRNAs regulate protein synthesis through the combined activities of two domains. The antisense region at 5' provides specificity to the target sense gene while the repetitive element confers the protein synthesis activation domain. This predicts that by swapping the overlapping sequence we may increase the amount of proteins encoded by the mRNAs of choice acting at post-transcriptional level. We thus synthesized a 72 nts long artificial sequence antisense to the AUG-containing region as transcribed from pEGFP. This sequence was inserted into AS Uchl1'5' to generate AS GFP. AS GFP strongly increased GFP protein levels in HEK cells when co-transfected with pEGFP, while it had no effects on its mRNA. When we pulsed cells with methionine for an hour and immunoprecipitated GFP, AS GFP induced an increase in radioactively labeled, neo-synthesized GFP, without affecting mRNA levels.

We then addressed the possibility that other SINEB2-containing lncRNAs may regulate the expression of their protein-coding partner through a post-transcriptional mechanism, based on similar structural elements. The FANTOM3 collection of non-coding cDNAs was bioinformatically screened for natural AS transcripts that contain SINEB2 elements of the B3 subclass in the correct orientation and 5' head to head overlapping to a protein coding gene. This identified 31 S/AS pairs similar to the Uchl1/AS Uchl1 structure. By sequence alignment, we were able to choose AS Uxt (4833404H03), antisense of Ubiquitously-expressed transcript (Uxt), as the one with the most similar SINEB2 elements. When AS Uxt was transfected in dopaminergic MN9D cells, it elicited an increase of Uxt protein level with no change in Uxt mRNA proving that AS Uxt, similarly to AS Uchl1, was able to increase protein levels post-transcriptionally.

Therefore we considered AS Uchl1 the representative member of a new functional class of lncRNAs named SINEUPs for their requirement of the inverted SINEB2 sequence to UP regulate translation in a gene-specific manner. The overlapping sequence is indicated as the Target Antisense Region while the embedded inverted SINEB2 element is the Protein synthesis Activation Domain. This work was published in Carrieri et al, Nature 2012 and it is the result of a collaborative contribution of SISSA, UCL and RIKEN.

It is common understanding that small and long non-coding RNAs are inhibitory of gene expression and translation, i. e. miRNA or RNAi. SINEUPs are the first example of a sequence-specific activator of translation representing a major achievement of DOPAMINET. The use of SINEUPs for protein manufacturing and RNA therapeutics has been patented and a START-UP company, TransSINE Technologies, have been founded by principal investigators of SISSA and RIKEN groups.

Global transcriptome of zebrafish dopaminergic neurons (ALU-FR, RIKEN).

Zebrafish develop dopaminergic neurons at anatomical locations correlated to most of the mammalian groups. However, the mes-diencephalic ascending

systems of groups A8-A10 in mammals have correspondence only in the diencephalon in zebrafish, where ascending systems originate from groups 1 and 3 of the posterior tuberculum. Distinct and prominent additional groups in zebrafish are; the ventral diencephalic groups 2 and 4, correlating with group A11 in mammals; the olfactory bulb DA group (A16); the hypothalamic groups corresponding to A12 and A14; and the pretectal group, which does not have a correlate in mature mammals.

As shown in D4.3 catecholaminergic neurons were labelled using transgenic zebrafish strains to drive expression of GFP. At stages 24, 36, 72 and 96 hrs. post fertilization, embryos were dissociated and GFP expressing cells sorted by FACS. Isolated RNAs were processed using either polyA selection and library generation or NanoCAGE.

Catecholaminergic neurons were labelled by four different strategies:
(1) 24 hrs. old embryos: we used the ETvmat2:GFP transgenic line which at this early stage labels catecholaminergic neurons in posterior tuberculum and locus coeruleus;
(2) 24 hrs. old embryos: we used Tg(otpb.A:egfp)zc48 transgenic line which at this stage label ventral diencephalic dopaminergic neurons and some preoptic neurons.
(3) For 72 and 96 hrs. old zebrafish larvae we used a th:GFP BAC transgenic lines that labels catecholaminergic neurons.
(4) for the 36 and 48 hrs. old zebrafish larvae we used a th:Gal4VP16 driver and UAS:EGFP responder transgenic line system to label catecholaminergic cells (Fernandes et al., 2012).

We used the different transgenic lines, because lines (3) and (4) do not efficiently label catecholaminergic neurons at early stages, while lines (1) and (2) also have GFP expression in several other non-catecholaminergic populations at later stages of development. Embryos were dissociated and catecholaminergic neurons were FACS sorted from GFP-tagged zebrafish (Manoli and Driever, 2012, Cold Spring Harbor Protoc.). RNA was either processed for NanoCAGE, or mRNA was isolated and amplified. cDNA was then sequenced by Illumina technique. This data submission is a series of data files consisting of three independent experiments with different RNA-Seq depth: Samples 1-4 (NanoCage): Samples 5-8 (RNA-Seq high read numbers), and Samples 9-12 (RNA-Seq low read numbers). The full set of data has been released to the public by submission to GEO (GSE41373).

This is a major contribution to the field since these are the first gene expression profiles of dopaminergic neurons in zebrafish.

The major use of these expression profiles in DOPAMINET has been for the identification of TFs expressed in dopaminergic neurons (see below).

Non-coding RNA transcriptome (ALU-FR, UoB, RIKEN, UCL, CBM).

To study lncRNAs, in D4.5 two complementary approaches have been implemented. 1. a bioinformatics analysis was chosen. Based on genome annotations available for Zv9.66, the genomic sequence around genes that are linked to dopaminergic development has been systematically screened to identify potential lncRNAs. 2. a Zebrafish transcriptome sequencing project was carried out at 2 cell, 30% epiboly, 14 somites and prim6 stage embryos while CAGE libraries were synthesized and sequenced at 12 stages (from fertilized egg to prim6). This was integrated with 3 stages of RNA sequencing and 2 stages of CAGE in Tetraodon. This work led to the

identification of over 1100 putative lncRNAs genome-wide (data generation of UoB, RIKEN, UCL in collaboration with B. Lenhard (ICL, London) and C. Nepal (U. Bergen)). As shown in D4.9, one lncRNA predicted by conservation analysis (collaboration with R Sanges, S Basu, Napoli) led to candidate lncRNA associated with onecut. This lncRNA is subjected to interference experiments (in progress).

Global transcriptome of Ciona dopaminergic neurons (CNRS).
Relatively little is known about the role and ontogeny of dopaminergic neurons in ascidians. Tyrosine hydroxylase expression is first detected at the late neurula stage in 2 precursors of the ventral central sensory vesicle, a territory thought to be homologous to the vertebrate hypothalamus. By the larval stage, TH expression is detected in 4-8 cells.

In D1.3-D1.8 CNRS first showed that electroporation of two reporter constructs (Ci-TH::GFP; Ci-Ptf1a::GFP) allowed the visualization of dopaminergic neuron precursors from the early tailbud stage onwards. However, several practical and theoretical issues suggested to change the model system to Fallusia mamillata: 1) FACS sorting was inefficient in Ciona due to the small fraction of cells to be sorted in each embryo (~4 dopaminergic neurons in an embryo counting greater than1000 cells), and the autofluorescence of several other cell types, 2) The number of cells that can be obtained was lower than anticipated. Ciona intestinalis adults usually produce less than 5000 embryos in their full reproductive season, a relatively large fraction (approximately 40%) of which develops abnormally after electroporation. Since the beginning of the project, animals from the usual source in Brittany (Roscoff Marine Biology station) had a shorter reproductive season than in previous years, so that the number of embryos was further reduced over most of the period, 3) high level of non-coding polymorphism between our local source of Ciona and the sequenced reference genome, making it difficult to align short reads in non-coding sequences. Importantly, it has been recently shown that a large fraction of genes are trans-spliced in Ciona (Matsumoto et al., 2010, Genome Research, 20(5):636-45), making promoter identification very difficult. In addition, we found that the currently known and used minimal promoters (pFog, pbrachyury) are sufficient for the analysis of distal cis-regulatory elements expressed in any Ciona territories, including dopaminergic precursors. This last element suggests that dopaminergic gene promoters do not include crucial information for dopaminergic precursor expression.

Phallusia mammillata produces several hundred thousands transparent non auto-fluorescent embryos, which develop with the same cell lineage as Ciona. We have thus sequenced, assembled and annotated the genome of this animal (see D3.6). Therefore, as shown in D2.3, we looked in Phallusia mammillata for the presence of cis-regulatory sequences of genes whose orthologs are expressed in dopaminergic neurones in Ciona. Because of the overall conservation of gene expression profiles between these two asciidans, these genes had a very high probability of also being expressed in Phallusia mammillata.

Analysis of gene expression in human (SISSA, RIKEN, UCL).
Although gene expression in human dopaminergic neurons was not among the initial aims of the project, we were indeed conscious that one of the long-term goal of DOPAMINET is the exploitation of its results for translational research in PD. Therefore, we integrated gene expression

data in mouse, zebrafish and Ciona with human datasets. To this purpose we used several strategies:
1. UCL has constructed a database of gene expression profiling in dopaminergic neurons including all published data from PD post-mortem brains. This database has been used in all our GRNs analysis (Taccioli et al, Database 2011).
2. We took advantage of FANTOM5 dataset from RIKEN. In this project, a simplified CAGE protocol adapted to single-molecule HeliScope sequencer (hCAGE) has been developed. hCAGE technology was used to profile over 1000 human and 500 mouse samples to build a promoter-level mammalian expression atlas and to model networks of distinct cellular states. Importantly, we were able to interrogate hCAGE datasets of over 60 brain libraries including human mesencephalon.
3. We studied gene expression in the blood of PD patients and controls (see below) to see commonalities and differences with dopaminergic neurons.
4. In D3.7 we also integrated AS TTS distribution with human FANTOM5 data collection. We focused our attention on antisense transcription to well-established human loci associated to Parkinson's disease. Antisense transcription was validated for a subset of genes, including a-synuclein, DJ-1, LRRK2 and MAPT. Most of the validated transcripts were predicted to have non-coding functions. (Zucchelli, Carninci, Gustincich, manuscript in preparation).

Overall, this DOPAMINET analysis represents the most comprehensive study to date of antisense transcription at loci associated to neurodegeneration and provides evidence for the existence of additional regulatory steps of disease-related genes by previously not-annotated lncRNAs.

With this work we have thus successfuly reached the following milestones:
M2: Microarray and Micro-CAGE analysis of pilot sets of neurons
M3: Complete set of mouse dopaminergic neurons isolated
M4: Microarray and Micro-CAGE analysis of complete sets of neurons.
The only deviation from the proposed plan concerned the expression profiling of Ciona. However, in the following pages we will show that by alternative experimental approaches we have substantially increased our knowledge on the TFs repertory of the 4 dopaminergic neurons per embryo of Ciona as well as we have been able to identify conserved cis-regulatory elements between Ciona and Fallusia.

A crucial objective of DOPAMINET was the "High throughput Screening of cis-regulatory elements"

Here we wanted to identify cis-regulatory elements and transcriptional networks that participate in the coding of dopaminergic neurons identity. To this purpose, we combined different approaches. First we aimed to screen candidate cis-regulatory elements in zebrafish and Ciona embryos, selected by intra-phylum phylogenetic footprinting, taking advantage of fast co-injection-based assays. By gene expression profiling and in situ hybridization experiments we listed combinations of TFs specifically present in selected dopaminergic subtypes. Finally, by identifying the promoterome of dopaminergic cells we searched for TFBSs enrichments. We can thus prioritize TFs to be tested in vivo with LOF and GOF experiments.

To analyze dopaminergic neuron-specific activity of putative Core Regulatory Modules (CRMs), KIT and UoB partners established a HTS

pipeline for the automated intelligent imaging of zebrafish embryonic brains at high resolution.

Development of new technologies: HTS in zebrafish (KIT, UoB).
The pipeline consisted of a novel protocol for sample handling and preparation and custom software modules for automatic identification and imaging of regions of interest.

The zebrafish embryonic brain is a bilateral symmetric organ thus its organization on the cellular and tissue level is best visualized using dorsal or ventral views, respectively. To achieve a standardized orientation of zebrafish embryos in wells of microtiter plates, templates for the generation of keel-shaped cavities in a thin layer of agarose were developed in close collaboration with engineering laboratories on the KIT campus. These cavities allow the standardized and tilt-free ventral orientation of zebrafish larvae facilitating the automatic acquisition of dorsal or ventral views on automated screening microscopes. During the course of this project 2 variants of the tool have been developed:
(i) a silicone based template for generation of grooves in agarose poured into a transparent tray and
(ii) a metal based template for the generation of cavities in standard 96 well plates compatible with chemical screening.

The silicone template for embryo orientation is easy to reproduce and has so far been distributed freely to several research laboratories in Europe and the United States conducting zebrafish imaging experiments that require standard orientation of specimen.

To visualize the cellular organization of the zebrafish embryonic dopaminergic system, high resolution multi-dimensional imaging has to be carried out. Standardly, researchers employ technologies such as confocal microscopy for high resolution three dimensional imaging. However, these technologies are usually limited in their imaging speed and thus not suitable for large scale high content screening approaches. In order to achieve the speed required for fast multi-dimensional image acquisition in combination with an image quality suitable for neuronal single cell imaging in oriented zebrafish, we utilized a standard wide-field screening system and expanded its functionality with a custom-developed pipeline for high-resolution, high-content screening. Custom-developed algorithms were implemented that can automatically detect regions of interest – such as the brain in this project – in low resolution pre-screen data. Then these coordinates can be extracted and the microscopic system can automatically image these regions of interest at higher resolution, enabling rapid capture of cellular resolution multidimensional data. This Matlab-based software toolbox provides the additional functionality to manually select regions of interest for subsequent high-resolution imaging. The software is freely available and downloadable from the lab homepage (see http://www.itg.kit.edu/liebel-lab-resources.php online). Moreover, a user-friendly graphical user interface has been developed allowing non-expert users to carry out intelligent high content screening approaches.

A major advantage of wide field microscopy is that imaging times are low allowing the rapid acquisition of multiple three dimensional datasets. However, widefield microscopy often necessitates the usage of image restoration techniques to remove blurred signal caused by intrinsic properties of the wide field setup. To achieve image restoration, we have

established a high throughput image optimization pipeline employing data handling scripts and batch deconvolution of z-stacks, which allowed us to perform higher speed and higher quality imaging at the same time. The computing power requirements of the deconvolution pipeline are quite substantial with 24-48 CPU cores for fast parallel processing. Thus, we developed microscope-compatible processing devices for fast data storage and "real-time" processing. Additionally, software modules for parallel processing have been established. The complexity of the images has an unpredictable influence on the time required for the image enhancement pipeline, therefore a special load balancing technology has been developed which automatically balances disk read/write operations and core activation.

The pipeline developed in this project was utilized to automatically image dorsal views of brain of the vmat2:gfp transgenic line in which monoaminergic neurons are labeled by GFP expression and were injected with putative CRMs linked to mCherry reporter gene. To test activity of CRMs in dopaminergic neurons an ImageJ based image processing workflow was established that allows the semi-automated generation of compiled expression patterns using maximum projection overlays and the automated analysis of colocalization within entire embryonic brains. Although initially motivated by the requirement to acquire and analyze high resolution datasets of zebrafish embryonic brains, the toolset image processing pipeline was developed such that it is highly flexible thus not limited to study zebrafish nervous system. By now, the tools and technology developed within DOPAMINET are being utilized in several additional projects in the field of tissue and whole organism high content screening. The results of the work conducted during the course of this project have been partially published in Peravali, R., Gehrig, J. et al. (2011) BioTechniques 50(5):319-324) and were editorially highlighted in Blow NS. (2011) BioTechniques. 50(5):275. Additionally, a detailed accompanying protocol for intelligent high content screening has been published in the BioTechniques 2012 Protocol Guide.

During the development of the image based screening technology for DOPAMINET a strong market potential was identified. As a consequence, the spin-off company "Acquifer" was founded in the beginning of 2012 (see http://www.acquifer.de online).

In D2.5, this HTS pipeline has been validated by screening for reporter activity of 202 enhancer-promoter combinations, based on images of thousands of embryos. (Gehirg et al, Nature methods 2009).

With this work we have thus successfuly reached the milestones M6: Zebrafish cis-regulatory element HTS screening system setup.

Development of new technologies: ViBE-Z (ALU-FR).
Precise three-dimensional (3D) mapping of a large number of gene expression patterns, neuronal types and connections to an anatomical reference is fundamental for reaching DOPAMINET goals in zebrafish. Therefore in D4.6 we developed the Virtual Brain Explorer (ViBE-Z), a software that automatically maps gene expression data with cellular resolution to a 3D standard larval zebrafish (Danio rerio) brain. ViBE-Z enhances the data quality through fusion and attenuation correction of multiple confocal microscope stacks per specimen and uses a fluorescent stain of cell nuclei for image registration. It automatically detects 14 predefined anatomical landmarks for aligning new data with the reference brain. ViBE-Z performs colocalization analysis in expression databases

for anatomical domains or subdomains defined by any specific pattern; here we demonstrate its utility for mapping neurons of the dopaminergic system. The ViBE-Z database, atlas and software are provided via a web interface (Ronneberg et al., Nature methods, 2012). This represents a major technological breakthrough by establishing a virtual 3D analysis and modelling framework that enables to map and analyze gene expression at cellular resolution in the context of dopaminergic and other neuronal system.

Transcriptional network in zebrafish (ALU-FR).
The obtained transcriptome data of zebrafish dopaminergic neurons have been analysed by CLC Genome Workbench to identify genes regulated at least 1.5-fold with statistical significance pless than0.05. Among them a series of TFs have been identified. As in D4.3, whole mount in situ hybridization confirmed that nkx2.1a and nkx2.1b, nhlh2, bsx, nr2e1, sox1a, zgc:153948 and zgc:171531 are indeed expressed in areas of dopaminergic neurons differentiation. All these TFs were either coexpressed in A11 neurons or co-expressed with otpa, which is required in precurors of A11-type dopaminergic neurons. We then extended the number of TFs included in this analysis by systematically screening gene expression databases (see http://www.zfin.org online) for TFs expressed in the proximity of ventral diencephalic dopaminergic neurons during zebrafish embryogenesis. This screening has identified a significant number of additional TFs, which were all subjected to coexpression analysis at the global anatomical level as well at cellular resolution.

An overview on transcription factor expression is given in specific dopaminergic neuron groups. The overview focuses on transcription factors that are informative with respect to regional identities in the brain. It also includes some previously published data. Importantly, we did not consider those transcription factors that transiently expressed, or expressed during neurogenesis before onset of TH expression, like ngn1. The data confirm previously published anatomical correlations between dopaminergic groups in zebrafish and mammals, and resolve several open issues with respect to group identities.

For the subpallial DA group, expression of dlx2a and dlx5a confirms that all neurons of this dopaminergic group, which extends along the subpallial/pallial border, actually reside within the subpallium. Thus, this group is indeed a restricted endostriatal dopaminergic system. etv1 expression in both OB and subpallial dopaminergic neurons suggests that an evolutionary ancient code for dopaminergic cells development is used in the telencephalon that dates back to nematodes.

The pretectal dopaminergic cells group is characterized by the transcription factors emx2, gata3, otx1b, otx2, and zic1, providing it with a molecular identity very distinct from other dopaminergic groups. zic1 defines the position of this group to be in the dorsal region of prosomere 3.

With respect to Otp-dependent dopaminergic neurons, an open question has been the origin of this group in the posterior tuberculum at the alar/basal plate boundary. In mammals, the Otp-dependent A11 dopaminergic neurons are located in the alar plate, but they co-express Nkx2.1, which suggested a migrated hypothalamic origin. The expression of nkx2.1b in subsets of DC2 and DC4, and of nkx2.1a in DC5/6 neurons support the common hypothalamic origin of all Otp-dependent dopaminergic groups. Further, we find lhx1a, lhx1b and lhx5 expressed in most of Otp-dependent

dopaminergic neurons, indicating that Lhx family genes may contribute to transcriptional specification of DC2 and DC4-6 neurons. Indeed, Morpholino knockdown of lhx5 has been shown to affect differentiation of Otp-dependent dopaminergic neurons.

nkx2.1a and nkx2.2a expression in group 3 dopaminergic neurons confirmed their anatomical assignment to the medial hypothalamus. DC3 is also the only dopaminergic group in our study expressing prox1. Interestingly, it has recently been demonstrated that morpholino knock-down of Prox1 results in a reduction of Otp-dependent dopaminergic neurons. Our finding that prox1a is expressed only in DC 3 but not in the Otp-dependent dopaminergic neurons (DC2, 4-6) suggests that Prox1a may act non-cell autonomously, or may only be expressed in precursors of these cells but not in mature Otp-dependent dopaminergic neurons. The expression of dlx5a in DC3 and DC7 hypothalamic dopaminergic neurons reveals that zebrafish hypothalamic groups share dlx5a as a potential transcriptional determinant with mammalian Dlx5 expressing arcuate nucleus dopaminergic neurons, and may thus relate to the A12 group.

We also asked whether there may be additional subgroups, or molecular heterogeneities among the dopaminergic groups initially defined by anatomical location. Most anatomically defined groups appear to coincide with group-specific common transcription factor signatures. However, we also observed expression of transcription factors in subsets of dopaminergic neurons within a group only, either separating domains, or appearing in a salt-and-pepper fashion. In some cases, this may represent true heterogeneity, as for example with arx-positive and arx-negative subpallial dopaminergic neurons. However, these findings may also be caused by the fact that dopaminergic neurons are continuously added to most of the groups during the developmental stages analysed. Therefore, fully matured neurons may be located next to TH expressing neurons that just exited the precursor population, and may still express precursor-specific transcription factors.

Does not provide any evidence for a core transcriptional code that would be common to all or most dopaminergic neuronal groups in zebrafish. However, the overview suggests that there are combinations of transcription factor families that specify several dopaminergic groups. These families include dlx genes, lhx/lim genes, nkx genes, pou3 class genes, and some nuclear orphan receptors in the diencephalon. In the telencephalon it is only etv1. On the other hand, genes like otp, sim1, gata3, otx or zic2a appear to be specific to dopaminergic groups in restricted brain regions only. Combining these two observations, we speculate that dopaminergic neuronal differentiation may be specified by a combination of region-specific members of a small, shared group of transcription factor families that may contribute to dopaminergic identity, in combination with selected region-specific factors that may attribute characteristic features (e.g. projection patterns) to each dopaminergic group.

In summary, we have made significant advances towards the understanding of the diversity of dopaminergic neuronal subtypes by having defined for the first time a complete molecular transcription factor code indicative of each major dopaminergic subtype in zebrafish.

Analysis cis-regulatory elements in zebrafish (UoB, FZK, RIKEN, UCL).

In order to test the functionality of cis-regulatory elements such as enhancers, core promoters are critical as they provide the platform for enhancers to communicate with transcription initiation complexes and provide enhancer promoter interaction specificity. As a key component of understanding the enhancer promoter interaction of neural genes, we mapped transcription initiation events in zebrafish embryos at single nucleotide resolution and generated a genome wide map of core promoters. In collaboration with RIKEN, CAGE and Illumina sequencing at 12 developmental stages were used to generate the first, high-resolution analysis of the promoterome during zebrafish development. The analysis uncovered the ontogenic stage dependent dynamics of core promoter architecture and utilization in genome scale and quantitative fashion. Besides protein coding genes, a wide range of developmentally regulated intra- and intergenic transcription initiation events have also been revealed. By complementing the zebrafish promoterome datasets with the one of the Tetraodon nigroviridis embryos, evolutionary conservation of transcription initiation regulation at over 160 million years of divergence has been identified. These comparative analyses supported the identification of novel transcription initiation mechanisms such as a non canonical AA-type initiator (AAI) used by membrane, vesicle and vesicle transport associated genes which are enriched in neural genes (Muller, Carninci, submitted).

Identification of candidate enhancers (UoB, FZK, UCL).
Recent years have seen the explosion of epigenetic data (primarily by the analysis of histone posttranslational modifications), which improved the annotation of functional elements of genomes. Among the histone marks, H3K4me1 and H3k27Ac are two modifications that appear to reliably predict cis-regulatory modules such as enhancers. We have carried out genome wide analysis of H3K4me3 (promoter) and H3K4me1 (enhancer) marks in zebrafish embryos (the latter in collaboration with UCL) in order to uncover putative promoter and enhancer sequences. In parallel, similar data were generated by other laboratories. Thus, several genome wide (published and our own unpublished) biochemical datasets become available for our work. As a first step towards the identification of dopaminergic neuron specific enhancers we have taken a gene list (generated by ALU-FR) which contained 28 transcription factors expressed at least in part in dopaminergic neurons in zebrafish larvae. We have postulated that transcription factors acting in the same sets of cell types are likely regulating each other and thus binding sites for several of the TFs are expected to be enriched in candidate enhancers active in the same sets of cells. The following TF clustering was used in subsequent TFBS clustering analysis: OTP GROUP (DC 2, 4, 5, 6) Arnt lhx1a, lhx5, nkx2.1, otp pbx1a, sim1 Hypothalamic: (DC 3, 7) Dlx5a pou3f3 pou3f1, nkx2.1a, nkx.2.2, prox1 and Rostral group: Arx, dlx2a, dlx5a, meis2.2, pbx1a, pou3f1/pou3f2. On the basis of this hypothesis and clustering, we have analysed TFBS distribution in candidate enhancers. Enhancer candidates were picked by two independent approaches.

Firstly, elements were chosen based on sequence conservation, which has previously been shown to be an indicator of functional significance. By comparing syntenic position between human and zebrafish we generated a set of candidate loci, which may contain regulatory regions relevant for candidate target genes. Next, we searched for candidate enhancers within these regions by looking for conserved elements between zebrafish and medaka. The rationale was that teleost conservation is expected to be broader and more comprehensive in detecting candidate enhancers. Thus 940 conserved non coding elements (CNE elements) with 70% identity, in 50 bp

window were found. TFBS enrichment was analysed by Cluster Buster. Thus 133 elements were selected from which 35 contained binding sites for TFs expressed in dopaminergic neurons. Then we looked for TFBS-containing elements which are expected to be active in the domain of expression of otp and which are expected to have overlap with activity of the ET Vmat2:GFP transgenic line (used as reference for enhancer screening). 123 elements were thus predicted. Before functional testing was carried out, we have evaluated several enhancer testing constructs and efficiency of transgenesis. We used Tol2 transposase based integration system and transient transgenesis as well as Sleeping Beauty transposase based approach. We have compared hsp70 and gata2 promoters and both appeared as reliable in interacting with candidate enhancers in pilot tests. Finally we have carried out enhancer synergy experiment by combining 3 elements in single constructs and compared their activity to single element containing constructs.

Among the 123 elements containing TFBS combinations reflecting the 3 expression sites (OTP expression domain, hypothalamus and rostral domain) 12 were functionally tested in a pilot screen using hsp70 promoter linked to RFP in the ET Vmat2:GFP line as reference. 6 of the elements showed neural enhancer activity and one of them with colocalisation with the ET Vmat2:GFP expression domain suggesting dopaminergic-specific activity.

The second approach ignored sequence conservation but used biochemical prediction (histone modification marks) followed by TFBS analysis using Cluster Buster. This approach led to approximately 70 elements (with some overlap to the elements described above). These elements were cloned into a gata2 promoter mCherry construct and injected into the Vmat :Gfp reference line. Work is still in progress, so far 35 elements have been functionally tested and a large proportion gave low but distinguishable neural activity. Among them 2 elements gave clear reproducible pattern of activity in telencephalic and hindbrain neurons.

Transcriptional network in Ciona (CNRS, UoB).
Prior to our analysis, cis-regulatory sequences for only two dopaminergic genes were known: TH (Moret et al., 2005, European Journal of Neuroscience, Vol. 21, pp. 3043–3055,) and Ptfb (Takeo Horie, personal communication). To identify Ciona TFs expressed in dopaminergic neuron precursors, we scanned the Ciona intestinalis expression data section of the ANISEED database for gene expressed at any time in the cell lineage that leads to dopaminergic neurons. This led to a first list of Ciona dopaminergic genes: AADC, Alas, Alox12, Cdc14a, Ci.R1CiGC29d23, DBH, Edem3, GHC1, Gonadoliberin, Gpm6a, Gyg88E, Meis, MSC, Msi1, Nkx2-4, Otx, Pacrg, pTF1a, pTFb, SERT, SoxB1, Spata17, TH, and TRIM36.

The Phallusia mammillata genome, which we sequenced, was blasted with protein sequences for each of these genes and in 5 cases, the corresponding Phallusia mammillata scaffold included a long 5' flanking sequence for the gene of interest: Otx, Meis, SoxB1, Alox12 and Agtr1a/b. For each of these genes, we PCR amplified and cloned a fragment of 3kb upstream of the genes of interest in an expression vector, upstream of the LacZ reporter plasmid. The resulting constructs were subsequently electroporated in both Ciona intestinalis and Phallusia mammillata. For each of the constructs tested, identical expression profiles recapitulating the history of expression of the gene of interest were observed in Ciona intestinalis and Phallusia mammillata. This shows that 1) in DOPAMINET we have identified 5 novel cis-regulatory regions from

ascidian dopaminergic genes, 2) the activity profile of these constructs is identical in the two species.

Genome-wide analysis of cis-regulatory elements.
We previously showed that, in Ciona and flies, cis-regulatory regions include a specific dinucleotide sequence signature associated to nucleosome depletion (Khoueiry et al., Current Biology, 2010). We however also showed that this signature, combined with evolutionary conservation, is not sufficient to efficiently identify cis-regulatory elements. To improve our predictions we analyzed two types of selected cis-regulatory elements in Ciona.

First, we generated a collection of synthetic elements derived from the early neural enhancer of Otx driving expression in dopaminergic neuron precursors. These elements preserved the sequence, position and orientation of crucial transcription factor binding sites, but randomized the intervening sequences. We tested 40 such randomized elements and found that their qualitative pattern of activity was indistinguishable from the parental natural element. By contrast their quantitative level of activity was highly dependent on intervening sequences, ranging from no activity to a level of activity significantly superior to the natural element. These experiments establish that intervening sequences do play a quantitative role in enhancer activity, and reveal that natural elements are not optimized for their quantitative level of activity.

Second, we computationally analyzed known cis-regulatory elements in Ciona, in search for some cis-regulatory signatures that we could experimentally test. The principle of the method was to integrate:
1) evolutionarily conserved transcription factor binding sites, using our extensive Ciona Selex dataset (currently 143/500 Ciona transcription factors were successfully analyzed by SELEX-seq, as a continuation of a previous EU project: Transcode);
2) levels of sequence conservation within the Ciona genus, and
3) regions predicted to be free/occupied with nucleosomes.

These predictions were run on a previously identified set of long flanking sequences with cis-regulatory activity in the hope of defining minimal cis-regulatory modules. Preliminary results on two genes suggests that the predictions can help defining minimal neurectodermal elements: Elk, expressed in the same cell lineage as Otx, and EphrinA-d, which has a broader expression in the ectoderm and neurectoderm.

Functional analysis of conserved shuffled elements (CBM, UCL, CNRS).

This consortium has also aimed for comparison of chordate cis-regulatory elements to elucidate evolutionary relationships in divergence of cis-regulatory mechanisms. Co-option of cis-regulatory modules have been suggested as a mechanism for the evolution of new expression sites during development. However, the extent and mechanisms involved in mobilization of cis-regulatory modules remains elusive. To trace the history of ancestral non-coding elements, which may represent candidates for ancestral cis-regulatory modules during chordate evolution we have searched for non syntenic conserved elements in olfactores genomes. We identified 183 non-coding sequences conserved localized in non-syntenic regions between vertebrates and tunicates which include Ciona species. These elements are associated with "transcription factor/developmental regulator (trans-dev)" genes, which are often characterized by enrichment for conserved enhancer sequences. The majority of these regions overlaps

ultraconserved elements in vertebrates and we carried out functional verification of the predicted chordate candidate enhancers. Using 3 Ciona candidate enhancer elements showing sequence similarity to zebrafish enhancers in heterologous genes we demonstrated that 2 of the 3 tested elements can both act as functional enhancers as well as can be transcribed in extant species of olfactores. The results support a hypothesis which argues for two possible shuffling mechanisms for the apparent non-syntenic preservation of such elements: genomics rearrengements followed by gene loss and retrotrasposition.

Besides chordate shuffling elements, additionally vertebrate specific enhancer shuffling was also interrogated. A set of highly conserved non genic elements were identified (CBM) which are duplicated in zebrafish and at least one copy is non-syntenic between zebrafish and mammals. 3 sets of such duplicated elements were tested for enhancer function by transient transgenesis using krt4 heterologous core promoter linked to venus GFP reporter and injected in transient transgenic embryos. For 2 sets both the orthologous and paralogous elements carried enhancer activity suggesting that enhancer cooption is a potential mechanism for generating new expression sites. (Sanges et al., Nucleic Acids Research 2013)

Development of new technologies.
To streamline and improve the throughput of ascidian electroporation assays, we designed a novel RNA-seq based reporter assay, whereby individual candidate cis-regulatory elements, placed in front of a minimal promoter, drive the synthesis of a barcoded reporter RNA. The aim is then to co-electroporate tens to hundreds of elements, simultaneously reverse transcribe, amplify and sequence the bar-code area of the reporter genes and thereby measure the relative level of activity of each individual construct. This system was validated using the Otx early neural element that drives expression in dopaminergic and other neurons. We showed in particular that individual 10-base barcode sequences did not influence the outcome of the assay. The assay is now ready to apply to a large scale candidate ascidian cis-regulatory elements.

With this work we have thus successfuly reached the milestones:

M5: Analysis of selected cis-regulatory elements in Ciona.
M12: HTS Analysis of 1,000 cis-regulatory elements in zebrafish. In Gehrig et al., Nature Methods 2009, 200 enhancer promoter interactions were analyzed. In Sanges et al., Nucleic Acid Res 2012, chordate conserved elements were verified for function in both fish and Ciona. In Nepal et al., submitted 2012, genome identification of tens of thousands of core promoters are described and 10 core promoters tested. In summary we achieved the analysis of 40 core promoters and over 70 candidate enhancers. Not reaching the 1000 elements were partly due to delay in computational identification of candidates and partly because far less candidate enhancer elements were computationally and biochemically (ChIP) found, which fulfilled our criteria for predicted dopaminergic specificity and as a result the original target has become untenable and unattainable.

Transcriptional network in mouse (SISSA, UCL, RIKEN).
In D3.5 by different bioinformatics strategies we identified the promotorome of DA cells, that is a list of promoters which use is enriched in A9 and A10 neurons. Promoters were defined as the genomic

region around the TC as in nano-CAGE dataset. Both CpG and non-CpG island-associated promoters were identified.

TCs were mapped to -500,+500bp region around Refseq TSS for each mouse gene. If multiple TCs were overlapping to the region around Refseq TSS, the one with the highest CAGE tpm was chosen as representative CAGE TC for that Refseq gene. The mapping was done separately for A9 and A10 CAGE data sets.

Following the TC mapping, differentially expressed TCs associated with Refseq genes were extracted for A9 and A10 cells. Differential expression was determined comparing CAGE expression values of A9/A10 TCs against OE, hippocampus, and cortex tissues and against each other (A9 vs. A10). This means we got sets of differentially regulated and gene-associated TCs for A9-A10, A9-OE, A9-Hippocampus and A9-Cortex pairs. We gathered up-regulated TCs and down-regulated and non-differentially expressed TCs in A9 cells compared to OE. We gathered similar set of TCs for A9-A10, A9-Hippocampus and A9-Cortex pairs. This provided sets of up-regulated TCs in A9 and sets of non-up-regulated TCs to be used as background when doing TFBS analysis. We extracted similar sets for A10 but this time we obtained up-regulated TCs in A10 cells compared to A9, OE, hippocampus and cortex cells.

When defining promoter regions, we used -400,+100 bp around the peak location (most used nanoCAGE TSS) of the Refseq associated TC. We have also distinguished between promoters that are CpG island associated and promoters that are not CpG island associated. We analyzed the CpG and non-CpG promoters separately.

In an alternative strategy the final set was obtained by taking all clusters that had normalized tag count in A9 larger than OE, Cortex and Hippo, and p.value lower than adjusted p.value cutoff. (0.05 adjusted by bonferroni method gave a cut-off of 1.86e-7). These dataset contained 1200 cluster subdivided in overlapping and non-overlapping CpG island (a cluster was designated CpG if there was an CpG island in +-1000 nt. from peak position): 1000 TC were associated to non-CpG and 200 to CpG.

Foreground sets were obtained by extending a region around peak positions by an arbitrary amount and filtered using one of three data sets:
1. H3K4me3 methylation
2. H3K4me2 regions with those that overlap H3K4me3 removed
3. PHASTCONS - multispecies conservation
Background sets were obtained by taking a random subsample from the set of non-differentially expressed clusters, but taking care that the normalized expression distribution matched the foreground set as much as possible. The background sets were then filtered using the above mentioned filters.

We then looked at the TFBS over-representation of differentially expressed promoters of A9 and of A10 cells. The idea behind the analysis is to find out the over-represented TFBS for up-regulated A9 genes and for up-regulated A10 genes. To this purpose, we used the list of up-regulated A9 and A10 promoters in D3.5. We did the over-representation analysis by comparing TFBS content of up-regulated promoters versus non-up-regulated promoters. In the final result overview, we included TFBS that were over-represented in all of the comparisons. We also analyzed CpG island-associated and non-CpG island-associated promoters separately.

TFBS analysis has been performed on 2 sets of enriched TCs as obtained by two independent strategies as described in D3.5.

TFBS analysis from D3.5 Strategy 1.
For the differentially expressed and Refseq associated TCs, we carried out sets of TFBS over-representation analyses using CLOVER software. The aim of the analysis was to find out over-represented TFBS in promoters of up-regulated A9 and A10 genes compared to various backgrounds. Two sets of analysis were performed, one for up-regulated promoters in A9 and one for up-regulated promoters in A10. To this aim, we have used four different background in each set of analyses.

When defining the promoter region we used -400,+100 bp around the peak location (most used CAGE TSS) of the Refseq-associated TC. We have also distinguished between promoters that are CpG island associated and promoters that are not CpG island associated. We analyzed the CpG and non-CpG promoters separately.

TFBSes over-represented in differentially expressed CpG promoters of A9 cells are depicted and non-CpG associated promoters. However, there were no over-represented TFBSes for A10 up-regulated and CpG associated promoters. But there were over-represented TFBSes for A10 up-regulated and non-CpG island promoters.

TFBS analysis from D3.5 Strategy 2.
Foreground and Background sets were scanned for TFBS, using PWMs from the whole Jaspar database.

The only set that had a substantial overrepresentation of a given TF was the d5000u10000 H3K4me2 filtered (regions were extended downstream 5000 nuc and upstream 10000 nuc from a given tag cluster peak position, and then filtered using a H3K4me2 filter with H3K4me3 regions removed).

As a third strategy we also took advantage of ChIP-Seq public available data on mouse cortical neurons of Transcriptional co-activator CBP and Histone Marks H3K4me1 and H3K4me3 (Histone H3, mono and try-methylated at lysine 4) marks in mouse cortical neurons (Kim et al, 2010), we carried out a broader analysis for cis-regulatory elements in genes expressed in dopaminergic neurons. Data for are presented in the file "3.6 summary enhancers complete" for Type 1, Enhancer, with signal found for H3K4me1 and CEBP but not for H3K4me3 or for Type 2, Promoter: for H3K4me1, CEBP and H3K4me3.

TFBSs that showed the greatest enrichment in strategy 1 were: EN1, Hand::Tcfe2a, Nr4a2, Stat3, Sox5.

Furthermore, according to strategy 2 analysis overrepresentation was found for En1, Nurr1, Pbx, GATA2, NHLH1 and HYF. (Lazarevic, Stupka, Carninci and Gustincich, manuscript in preparation)

This was very interesting since En1 and Nr4a2 are two of the most important TFs involved in mesencephalic neurons differentiation and maintenance. These results represent an important independent validation of our experimental and bioinformatics approach. Furthermore, it is the first evidence to date that different repertory of TFBSs are enriched in promoters as identified by TSS distribution in A9 and A10 neurons proving that different GRNs are present in the two subtypes.

With this work we have fulfilled Milestone M9: A list of core dopaminergic neurons core promoters, non-coding RNA candidates TSSs, TFs and TFBS expression and Sense/AntiSense status.

Systems biology can be characterized as an integrative experimental-computational approach where mathematical modelling is used in order to gain insight into the underlying biological mechanisms of a particular process or disease. As defined in the objective "Network Reconstruction" our aim in DOPAMINET is to identify the gene regulatory networks responsible for the definition of single dopaminergic neurons. Such an understanding may capture in computational models neuron-relevant processes and thereby pave the way for identification of putative nodes within the network as well as "master switches" within each network. In particular, we aimed to assess whether dopaminergic neuron networks are comparable across chordate organisms and what are the differences between subtypes.

To answer these questions YH and UCL first generated a database, named ParkDB, that contained all the datasets considered. This was accomplished to collect all the data published on dopaminergic neurons and PD and integrated them with DOPAMINET datasets available at the time. This allowed the identification of key differentially expressed genes. Furthermore, it provided homogeneous results for further analysis by the scientific community at large. First, we assumed that the available datasets associated to Dopaminergic Production allows the identification of relevant genes. For each data set we defined a list of Differential Expressed Genes. This allowed the definition of a distance between datasets, which is easily done if datasets are of the same biological species, but needs to rely on orthology if they are not. The intersection between genes of different species was done by mapping genes in dataset 2 (da2) to genes in dataset 1 (da1). By using this approach we developed a distance between every pair of datasets. The distance had a range between 0 and 1, and was used to represent the dataset in 2 or 3 dimensions. In our case (and in addition to the classical heatmaps) we considered Multi-Dimensional Scaling (MDS) for this visualization. The purpose of MDS is to transform measures of similarity or preference (eg. Distances between genes; or in our % of agreement in differential expression) into distances represented in multidimensional space. The resulting perceptual maps showed the relative positioning of all objects (datasets). We made this comparison more generic by including different measures or selection: adjusted p-value, p-value or Fold-change to compute the percentages in agreement. In addition we considered computing the non-parametric correlation between those values in all genes for all pairs of datasets.

Interestingly, Spearman and DEcomp provided different results in association of datasets with DEcomp providing more coherent results with the biological description of the datasets.

After moving to a bi-dimensional space by MDS, results for Spearman and DEcomp showed that comparisons are primarily grouped by Species. By this observations we may conclude that "overall" there is no evidence of a shared mechanism among the different gene networks responsible for the dopaminergic phenotype among the different species. This answers one of the major questions addressed in DOPAMINET.

However, we were able to identify a set of genes associated to the core of the comparisons under consideration. Interestingly, zebrafish data

seems to be at the center of the MDS suggesting it may contain the very central core-model of dopamine production. As an alternative interpretation zebrafish may be far from the other species being in the middle since its distance is equally far from all of them. Despite these results we identified a of genes as at the core of the comparisons. This list includes: Cd52 Clec14a St8sia6 Foxi1 Hapln4 Zfp61 Galntl2 A930005I04Rik Tnnc1 Plekhm3 Lbxcor1 Mettl7a1 Stk3 Rps17 Twf2 Ttr Il23a 3110035E14Rik Rps20 Zfyve19 Rps6 Rps13 Hdac1 Ifna1 Slc30a2 Ifna11 Calca Zfp109 Il2rg 2610036L11Rik Bmper Aarsd1 Tbx18 1700113I22Rik Ano8 Ly6a Wdr18 2410002O22Rik Sstr4 Ly6c1 and AI894139.

Therefore our study suggested that we need to study each species independently and then compare between the different networks identified. These results have been published in the DOPAMINET paper: Taccioli et al, Database 2011.

We then focused our attention on the generation of a regulatory network of the mouse since this is the chordate that we have the most complete dataset.

Here we proceded by 1. Identification of Identification of Differential Expression Genes (DEG) for each data set; selection of genes which appeared to be DEG in most datasets; we name it DEGselect; considering DEGselect, for each dataset construct a network based on c3net approach; generate a consensus network; study the network generated.

The network thus obtained contained 1046 nodes and 15106 undirected edges. The undirected network showed a highly associated group of genes. This network was fairly complex so we created Network 2, which made a more conservative selection of links. In this network threshold was set greater than0.2. Here the number of nodes was 83. One of the most interesting observations is that the most relevant sub-network, and the group of nodes associated are unknown genes. For instance the gene with highest centrality (ENSMUSG00000071613, Gm10340) is a predicted protein whose function is yet to be determined. This may be said for the rest of the genes. In addition a second cluster is associated to Histone proteins, which would provide a hint on the epigenetic regulation of the system. A third cluster is associated to ribosome and nuclear regulation. Importantly, many of the genes known to be associated to Dopamine production and PD are certainly part of the network.

The identification of these unknown genes is an important and unexpected contribution of DOPAMINET suggesting the need for experimental studies of their function in dopaminergic neurons.

In the objective "Network Perturbations and target identification" we propose to validate by Gain-of-Function (GOF) or Loss-of-function (LOF) experiments the role of candidate TFs identified in DOPAMINET.

To do so we took advantage of three experimental settings.

Perturbation experiments in zebrafish (ALU-FR).
We have previously defined a set of transcription factors expressed in selected dopaminergic neurons in zebrafish. Here we performed individual and combinatorial LOF experiments by in vivo knockdown using antisense Morpholino technology. For those TFs that showed a specific DA loss-of-function phenotype, we performed GOF studies by Heat-shock pomoter driven

overexpression of the TF in the embryo and early larvae, and analysis of effect on ventral diencephalic (vDC) DA development.

Several Nkx2 TF family members are expressed in the vDC region of dopaminergic cells differentiation. We have previously shown that the only zebrafish DA group with ascending projections, A11, express nkx2.1a and nkx2.1b and that nkx2.1 is also expressed in mouse A11 DA neurons. Therefore, we wanted to determine whether the three zebrafish Nkx2 family TF may act in a partially redundant fashion in dopaminergic differentiation. We thus performed individual knock-downs which did not cause severe abnormalities in DA development. Furthermore, any combination of double knockdowns had only mild effects on vDC DA specification. In contrast, when we performed a triple knockdown of nkx2.1a, nkx2.1b and nkx2.4, we detected complete loss of all vDC DA neurons, including the ascending DA groups. Our data establish Nkx2 family members as crucial contributors to ventral DA specification in zebrafish (Manoli and Driever, manuscript in preparation).

We then knocked-down Arx expression by antisense morpholino microinjections and analysed DA neuron development in morphant larvae by in situ hybridization to th and anti-TH immunohistochemistry at 72 and 96 hpf. We observed a significant decrease in the number of both prethalamic and preoptic DA group neurons. Thus, Arx appears to be essential for proper expression of the DA neurotransmitter phenotype in the prethalamus.

Functional inhibition of Isl1 by LOF experiments resulted in a significant decrease of the expression in the prethalamus, as assessed by in situ hybridization and immunohistochemistry at 96 hpf (Data published in the DOPAMINET manuscript Filippi et al., Developmental biology 2012). The analysis of slc6a3/dat and ddc expression also showed a reduction of their prethalamic domains, arguing for a role of isl1 in the formation of group 1 DA neurons. This result was confirmed by the analysis of th expression in the isl1sa0029 mutant line, which also displayed a reduction of prethalamic DA neurons.

Interestingly, following overexpression of sim1a or heat shocks of nkx2.1 alone, no additional DA neurons have been induced. In contrast, combined GOF studies of Sim1a and Nkx2.1 revealed a significant increase in number of DA neurons.

We selected nkx.2 family genes as well as otp, isl1 and arx for epistatic analysis to be carried out with LOF and GOF experiments. Interestingly, Nkx2.x triple knockdown embryos are deficient in expression domains of shh in the hypothalamus, and of otpa and lhx6 in the vDC DA precursor territory. Thus, in accordance with genetic studies in mice, nkx2.x genes have an early patterning function in the vDC. During previous work we have already characterized downstream targets of otp, which has a crucial function in vDC DA development in zebrafish. In the context of DOPAMINET, we identified a melanopsin, opn4a, which is expressed downstream of Otp in dopaminergic group 2 vDC DA neurons as well as in adjacent preoptic region light sensing deep brain photosensors. This works establishes that downstream of Otp multiple aspects of dopaminergic differentiation are controlled, and for the first time also established that vDC DA neurons are potentially light-sensing neurons. (Fernades et al. Current Biology 2012).

Furthermore, our results suggest that the inhibition of Arx activity leads to an impairment of dorso/ventral patterning of the prethalamic domain and loss of prethalamic DA and GABAergic neurons. The role of Isl1 in prethalamic DA differentiation appears independent of Arx.

As a major achievement of DOPAMINET, we have thus identified a significant number of epistatic relationships that place crucial TFs into a regulatory hierarchy during ventral diencephalic DA differentiation.

With this work we have thus successfuly reached the milestones M13: Technology development for LOF functional interference with a small set of ncRNA and TF in fish. We indeed concentrated our efforts on the analysis of TFs since we had a long list of promising candidates.

Perturbation experiments in the dopaminergic cell line MN9D (UCL, YH, SISSA, RIKEN).
Since TFBS for Nurr1 was the most enriched motif associated to TSS identified in mouse dopaminergic cells as in D3.4, D3.8, we carried out perturbation experiments by inducible expression of Nurr1 in the dopaminergic-like MN9D cell line. Experiments were conducted at SISSA, where RNA was extracted at several time points (12, 24, 36, 48, 60, 72 and 96 hours) in triplicate, to produce a detailed time series analysis of dopaminergic neuron differentiation. This model was chosen because it was an amenable model for a pilot analysis allowing us to obtain time series information which would not be easily accessible from an in vivo system.

At UCL the gene expression data was used to create an unsupervised plot showing a clear separation of time groups, and in particular an overall switch pre and post 48 hours. Moreover, positive controls (i.e. genes which are known to be regulated by Nurr1) were verified in the microarray data, indicating clearly that they were found to be upregulated as expected. Having established the validity of the experiment we proceeded with the differential expression analysis. This was performed with the time-course specific program maSigPro, an inferential methodology based on regression. Briefly, maSigPro fits a regression model by considering time as a continuous variable and creating specific variables for each treatment, thereby adjusting a temporal profile for each time point. The method assesses the significance of the global model (i.e., if there are significant differences with respect to time) and of each variable (i.e., which specific time change is present). The top differentially expressed genes are listed, and include well-known genes related to dopaminergic biology such as synuclein. Using a clustering approach we were able to classify genes based on their expression trends over the time course in 9 clusters.

To carry out a network analysis we took advantage of ARACNE (an information-theoretic algorithm for the reverse engineering of transcriptional networks from microarray data). This approach allows the identification of major functional groups and provides a hypothesis on the time-dependence between them. First subsection details the clustering approaches considered; second subsection details the network approach and a finally the third subsection integrates the reverse-engineered network with other network approaches that prioritize the identification of TF association. In the analysis we considered greater than2000 probes to be differentially expressed.

We run ARACNE over the set of differentially expressed genes. We considered several p-value thresholds to define significant relations: 10e-7,10e-10 and 10e-12. We selected the second option as it provides a good trade-off between the number of genes and the number of relations.

The major observation is the existence of two major clusters of genes. We observed that C3 is associated to inhibition of neuron-associated genes while C5 is associated to expression of mRNA-machinery genes (including chromatin transformation).

To investigate the TF regulation we made use of DREM 2.0 (Schulz et al,2012), which allows the identification of dynamic regulatory networks from time-series data. The idea is that under the assumption that we know the list of TFs and its associated targets the method integrates times series and static data using an Input-Output Hidden Markov Model (IOHMM). DREM learns a dynamic GRN by identifying bifurcation points, places in the time series where a group of co-expressed genes begins to diverge. These points are annotated with the TFs controlling the split leading to a combined dynamic model. Interestingly the TFs associated to the selected clusters were NKX3-1, PPARA and NFE2L1.

The very same RNAs preparations were also used to synthesize nanoCAGE libraries as reported in D1.12. After sequencing, an average of 8.8 million tags per sample was extracted and mapping rate reached 48% when tags were mapped to multiple positions, which are enriched for expressed repeat elements. 40 to 60% of the mapped tags were associated to known coding genes and the remaining reads represent potential matches to non-coding genes.

The smallest Spearman correlation coefficient between adjacent time points was 0.6. The Motif Analysis Response Activity method (MARA) allowed us to integrate the promoter location, presence of transcription factor binding site motifs and promoter expression levels, to determine potential regulators. We observed that immediate early transcription factors have a close correlation between gene and motif activity within the same time point. We also found that significant changes happened in the gene network of MN9D cells after 48 hours upon Nurr1 activation. Moreover, we also observed long-lasting effects that are potential cases of regulatory relays.

Perturbation experiments in primary dermal fibroblasts (SISSA).
Seminal studies have demonstrated that functional neurons can be generated independently of stem cells by direct cell conversion through genetics-based approaches. Therefore, in a collaboration with Dr Vania Broccoli at San Raffaele in Milano, Italy, we aimed to generate dopaminergic neurons through the direct conversion of somatic cells by forced expression of lineage-specific factors including Nurr1. As presented in Caiazzo et al., (2011) Nature 476 224-7, Nurr1, in combination with Mash1 and Lmx1a, were able to generate directly functional dopaminergic neurons from mouse and human fibroblasts without reverting to a progenitor cell stage. Induced dopaminergic (iDA) cells released dopamine and showed spontaneous electrical activity organized in regular spikes consistent with the pacemaker activity featured by brain dopaminergic neurons. The three factors were able to elicit dopaminergic neuronal conversion in prenatal and adult fibroblasts from healthy donors and Parkinson's disease patients. Microarrays analysis carried out in SISSA laboratory showed that iDA cells clustered with A9 and A10 adult mesencephalic dopaminergic (mDA) neurons (as shown in D1.9) rather than

with their fibroblasts of origin as illustrated by hierarchical clustering and the general degree of gene expression overlap. Of note, many representative genes of the dopaminergic phenotype, such as Th, Vmat2, Aadc (also known as Ddc), Ret, Gfra1, Foxa1, Gdnf and Drd2 were highly enriched.

We then took advantage of the cell conversion assay to test the activity of TFs found expressed in dopaminergic neurons and which TFBS was enriched around TSS obtained with nanoCAGE as shown in D3.8. Full-length clones for LMO2, LMO3, Etv5, HIF, PBX3 GATA2, LRXb, NeuroD6 and Nhlh1 were introduced into a lentiviral vector and tested in combination with Mash1 and Lmx1 for their ability to increase fibroblasts differentiation into dopaminergic neurons. Importantly, Etv5, HIF and PBX3 strongly increased the number of TH-positive cells. We are currently gene expression profiling these cells to monitor whether they resemble a specific mesencephalic dopaminergic cell type.

These results demonstrate the validity of the DOPAMINET approach to identify TFs able to trigger dopaminergic cells' differentiation.

With this work we have thus successfuly reached the milestones:
M10: Reconstruction of dopaminergic neuron networks in several organisms and neurons.
M11: Microarray and Micro-CAGE analysis of perturbation experiments
We plan to repeat some of the analysis for M10 with the most recent gene expression dataset obtained in zebrafish.

Analysis of gene expression in PD patients (SISSA, RIKEN).
A general problem with neurological diseases is that the site of degeneration is not accessible for direct study during life. Most importantly, these disorders are characterized by a long pre-symptomatic phase, lasting several years, during which degeneration is occurring but no clinical symptoms are evident. It is therefore clear that a pre-symptomatic diagnosis may allow potential drugs to act longer on a larger number of less compromised cells. Gene expression profiles represent an innovative tool to discover biomarkers. The challenge is to identify gene expression signatures as candidate biomarkers for PD and analyze their potential for early diagnosis in pre-symptomatic patients and their efficacy in clinical trials of new therapeutic treatments.

We were also very interested in assessing whether changes in the blood was mimicking gene expression in mesencephalic dopaminergic neurons.

We thus carried out nanoCAGE analysis and Affymetrix gene expression profiling of blood from 40 de novo PD patients and 20 controls. PD patients (age 64±8.4) have been diagnosed by neurologist board-certified movement disorders specialists at the Neurological Clinic, Ospedale of Cattinara, Trieste (Prof. Pizzolato) and met clinical diagnostic criteria for probable PD. We focused on de novo (drug untreated) patients that were evaluated by a specialist for the first time and who had no specific pharmacological treatment prior to evaluation. RNA was then collected from whole blood of 40 de novo PD patients and 20 controls. Affymetrix gene expression experiments were performed. The very same samples have been used for transcriptome analysis with nanoCAGE. We have sequenced on average 1.5M tags from each library.

Gene expression profiling was able to discriminate PD patients from age-matched controls. Significant changes were identified by analyzing both

mean expression levels and preservation of co-expression relationships. The majority of genes differentially expressed in blood are also present in mouse A9 neurons and human Substantia Nigra, the site of neurodegeneration. Interestingly, Gene Set Enrichment Analysis and Gene Ontology revealed chromatin remodeling and methylation as biological functions significantly altered in the disease. Candidate transcripts were validated by qRT-PCR. (Calligaris, Carninci and Gustincich, submitted).

We have been extending our analysis to genetic PD patients. To this purpose we harvested blood from 12 patients with parkin mutations (as an example of recessive cases) and 12 with LRRK2 mutations (as an examples of dominant mutations). Furthermore, we carried out a follow-up study analysing gene expression profiling of the very same patients but after one year of levo-dopa. Bioinformatic data analysis is in progress.

The identification of a gene signature for PD diagnosis led Gustincich to win the first prize at the Friuli-Venezia Giulia Start Cup competition for innovation. Gustincich then won the Italian National prize for Innovation (Working Capital, Torino 2011); 1st out of 2500 participants.

Gustincich, together with Carninci (RIKEN), then founded PARKscreen, a start-up to develop genomics-based diagnostics for neurodegenerative diseases.

Potential Impact:
DOPAMINET project presented a highly innovative and interdisciplinary approach to a disease, Parkinson's (PD), that has a deep impact in modern society due to its prevalence and its target population.

PD is an age-related common degenerative disorder that affects more than 1.2 million European citizens today. It is the second most common progressive neurodegenerative disease sickening 1-2% of all individuals, men and women equally, above the age of 65.

Although the project aimed at shedding light on fundamental aspects of dopaminergic neurons, such as the fine molecular networks at play, the potential impact of this knowledge in PD is clearly significant.

First and foremost mesencephalic dopaminergic cells are the primary site of neurodegeneration in PD. Therefore, the knowledge produced during the time frame of the project led to a better understanding of the genes expressed in dopaminergic neurons.

Among the most far-reaching discoveries, we list the following:
We now know the entire repertory of channels and receptors of these cells. These are under intense scrutiny in PD and are classic drug targets.
The identification of a new class of molecules, the olfactory receptors, as selectively expressed in these cells suggesting that specific types of odorant-like molecules may act as psicoactive drugs on dopaminergic neurons.
The discovery of new isoforms of alpha-synuclein may let us develop new drugs to induce or inhibit the synthesis of the c-terminal part of the protein.
The identification of transcripts antisense (AS) to genes involved in hereditary PD like AS to LRKK2, alpha-synuclein, UCHL1, DJ-1 and MAPT may provide new regulatory molecules to manipulate expression of genes involved in the disease.
The identification of the repertory of Transcription Factors that are differentially expressed between A9 and A10 neurons.
The identification of the Transcription Factor Binding Sites that are differentially enriched between A9 and A10 neurons.
The discovery of a combinatorial code in zebrafish for dopaminergic identity.
The study of gene expression profiling in the blood of PD patients can have an impact in our ability to diagnosize the disease and to follow the effects in vivo of new therapeutic treatments during clinical trials.

Most importantly, the identification of new gene networks involved in dopaminergic cells' differentiation and maintenance has allowed us to create cocktails of transcription factors able to trigger cell conversion from a dermal fibroblast to a spiking, dopamine-releasing neuronal cells. This will have a profound effect in future strategies for restorative therapy of PD. By over-expressing cocktails of A9-specific Transcription Factors we may trigger the differentiation of the very same dopaminergic neuron that degenerates in disease.

Therefore, the knowledge produced in DOPAMINET led to a better understanding of the disease, but more importantly has the potential to promote better health and quality of life and to reduce the high health care costs related to the treatment and care of elderly cognitively impaired patients.

Implications of the project results can be summarized in three major points as follows.

Improving European Public Health. PD is a progressive neurodegenerative disease that affects more than 1.2 million Europeans. This number is forecast to double by 2030. Although PD is most common in the over 60's, many people are diagnosed in their 40's and younger. Prevalence is expected to grow sizably over the next years as the proportion of aging population continues to increase. Europe has a rapidly ageing population, many suffer from PD and, according to the data above, the burden placed by Parkinson on the working-age population will rise dramatically. This is a challenge for the European society. Thus, any treatment that can impact on PD may have profound effect on European health.

Improving European Economy. The total cost of caring for patients with PD is huge. The annual European cost of the disease is estimated at 13.9 billion EUROS, and as our population continues to live longer, this cost will continue to rise dramatically – especially in the later stages of the disease where the impact is greatest on people with PD, their families and carers, and society as a whole. Even a modest improvement in the prevention and treatment of PD will therefore have economic impact, in particular in an ageing population like the Europeans. A second potential economic benefit is for European industry. The identification of genes that play a fundamental role in dopaminergic function provides a list of potential drug target for therapeutic treatments. Patents have been filed and START-UPs have been founded (see below).

Strengthening European Research. The joint effort of the consortium undertaken for DOPAMINET allowed to carry out an advanced project, to exchange know-how, experience and technologies, and thereby to tackle the research challenge using novel strategies with a highly competent team. It is through such collaborations that Europe can compete in the international world of science, thereby providing cutting edge research at the same level as that obtained by the large institutions in US and other places.

Essential pre-requisites for impact are dissemination and exploitation of the results.

Dissemination refers to the processes involved in getting the right information in the right format to the right people at the right time. This is of a particular relevance in a project like DOPAMINET where a diverse group of stakeholders are involved: scientists, students, journalists, biotech and pharma companies, doctors, patients associations and the public.
Original scientific data must be disseminated effectively in the scientific community with seminars and reports at conferences.

Basic research data and their implications for PD must be accurately communicated and discussed with patients associations and the general public.

On the other hand the experience accumulated in DOPAMINET may be capitalized in the future by policy makers and industries to better design future lines of intervention in basic research in molecular neuroscience and its translation into the quest to find a cure for this devastating disease.

Dissemination was carried out according to the type of audience we targeted and addressing the following objectives:
1) to inform and disseminate the work in progress as well as final results to the scientific community and provide it with training opportunities on the project activities and topics;
2) to raise awareness and sensitivity on the topics dealt within the project among non-specific public;
3) to promote commercial exploitation of project results.

A project website, consisting of a restricted and a public session, was set up and updated (see http://www.dopaminet.eu online). Promotion and communication activities submitting press-releases, publishing web-news across the social media and internet groupings were also performed.

Scientists-to-scientists dissemination occurred through several channels: publication in highly prestigious journals (more than 40 papers were published in highly relevant journals such as Nature, BioTechniques, Bioessays, Nature Methods, Nature Communications, BioTechniques, DNA Research, Experimental Medicine, Bioinformatics, Database the Journal of Biological Databases and Curation, Developmental Cell, Neurogenetics, PLoS, European Journal of Neuroscience, Genome Research), seminars and conferences. Join meetings with other EU-sponsored consortia such as mdDAneurodev were organized (Freiburg, Germany). Dissemination of DOPAMINET data has been very successful within the scientific community for the high quality of scientific achievements.

Three summer schools on Dopaminergic Neurons were attended, introducing the genomics of PD to a large number of young scientists of many countries. Internationally renowned speakers including beneficiaries of other relevant European projects were invited to these training events, open to researchers of the field from across Europe. Summer school 2012, in particular, was attended by about 100 participants from 15 nationalities (Austria, Czech, Republic, Slovenia, Serbia, Croatia, France, Italy, Japan, Vietnam, Germany, Israel, Palestine, Poland, UK and Russia). We were particularly resourceful in attracting female PhD students and post-docs to give a contribution in filling the gender gap.

Special emphasis was put on reaching out to the PD patients' associations. To this purpose, efforts have been done to increase communication skills of basic researchers to medical doctors and to the public at large. In particular, a workshop on scientific communication was designed and held in order to teach scientists how to effectively communicate scientific results and ideas to the public as well as to policy makers. In general DOPAMINET has provided an important contribution to make the general public understand of the fundamental roles of basic research to find a cure for PD.

Periodic outreaching events were held with Charity organizations, among them in Freiburg with German Parkinson's Foundation and in London with Parkinson's UK.

A roundtable on "Neurodegenerative diseases: from bench to bedside. Institutional strategies and role of the private sector in the transfer of knowledge" was organized in Trieste, Italy, to discuss the impact of DOPAMINET on science, health care and future EU-sponsored programs. The roundtable was the closing session the DOPAMINET 2012 Summer School and represented the chance to extend the discussion on the implications

of neurodegenerative diseases to society and to other stakeholders, such as policy makers, clinicians and innovative and research-intensive small and medium-sized entreprises (SME).

Invited panelists shared their experience on tools and strategies for increasing the engagement of industries and to develop an aggressive and innovative agenda for accelerating research and drug development for neurodegenerative diseases. Among policy-makers, confirmed speakers included prof. Adriana Maggi, Vice-Chair of the Joint Programming Initiative on Neurodegenerative Diseases (JPND).

The roundtable was attended by an hundred of participants and, in a common consensus, identified different aspects of neurodegenerative diseases that urgently need an implementation, such as:
- a better molecular definition of the different stages of the diseases and of their heterogeneity
- pre-symptomatic diagnosis
- the construction of a biobank with standardized procedures for sample collection and an open IT infrastructure
- larger private and public financing of basic research
- the creation of a new generation of research-trained clinicians
- home-care of patients through the help of an integrated IT structure
- need of efforts to increase awareness of these diseases at social and political levels.

DOPAMINET results are encouraging and may be relevant for therapeutic and translational approaches. To these aims DOPAMINET beneficiaries have given special attention to the technological transfer of the data. During the time frame of the project many technologies and two patents were developed and new START-UPs established.

RIKEN and SISSA developed NanoCAGE, a technology that miniaturizes the requirement of CAGE for starting RNA quantity to the nanogram range and which can also be applied to RNA obtained from fixed tissues (Plessy et al. Nature Methods, 2010).

This led to the Patent PCT -Patent filing No.PCT/JP2009/061552 Date: 2009/6/18 OWNER RIKEN, a method of manufacturing a mixture of amplified double-stranded nucleic acids comprising unknown sequence semi-suppressive PCR for nanoCAGE. The patent can be applied in the biotechnology domain and will be exploited via licencing.

From the gene expression profiling analysis of blood of drug-naïve PD patients, a gene signature to diagnosize PD has been identified. This led Gustincich to win the first prize at the Friuli-Venezia Giulia Start Cup competition for innovation. Gustincich also won the Italian National prize for Innovation (Working Capital, Torino 2011, 1st out of 2500 participants). Gustincich (SISSA), together with Carninci (RIKEN), then founded PARKscreen, a START-UP to develop genomics-based diagnostics for neurodegenerative diseases.

During the development of the image based screening technology for DOPAMINET a strong market potential was identified. As a consequence, the spin-off company "Acquifer" was founded at the beginning of 2012 (see http://www.acquifer.de online). Two of the former Dopaminet fellows play an active role in Acquifer (JG and UL). KIT and EMBL (European Molecular Biology laboratory) joined Acquifer as shareholders. Acquifer provides the entire pipeline of screening technologies as a service for full grown

platforms. This includes custom microscope development, image processing routines and cloud based collaboration services.

The identification of SINEUPs as a new functional class of antisense lncRNAs has opened a new field of research. Recent results from SISSA and RIKEN suggest that SINEUPs are widely distributed from plants to human and participate in translational control during stress. Manipulation of SINEUPs sequences allows in principle the design of a synthetic molecule able to increase protein synthesis at will and SINEUPs may be ideal therapeutic molecules for increasing protein levels in haploinsufficiencies. This led to Patent PCT "Patent filing No: US SN 61/469399 Date: 2011/3/30; Patent filing No.PCT/JP2012/059430 Date: 2012/3/30" – JOINT ONWERSHIP SISSA/RIKEN. A Method to enhance Protein Production Functional Nucleic Acid Molecule And Thereof.
Carninci and Gustincich thus founded TransSINE technologies (see http://www.transsine.com online), a new RIKEN venture to commercially exploit SINEUPs molecules.

List of Websites:

http://www.dopaminet.eu