Scuola Internazionale Superiore di Studi Avanzati

PhD Course in Functional and Structural Genomics

Discovery of a human VNTR allelic variant in Nprl3 gene intron that enhances its transcription in peripheral blood

Thesis submitted for the degree of "*Philosophiæ Doctor*"

Candidate
Maria Bertuzzi

Supervisor
Prof. Stefano Gustincich

Academic Year 2014-2015

Table of contents

Abstract

Parkinson's disease (PD) is a slowly progressive degenerative disorder of the central nervous system that is classically defined in terms of motor symptoms consequent to degeneration of specific subsets of mesencephalic dopaminergic (DA) cells within *substantia nigra* (SN) *pars compacta*. No pharmacological treatment is currently available to slow or arrest the neurodegenerative process. Furthermore, accurate early diagnosis suffers from the lack of reliable biomarkers.

By the time motor symptoms appear, PD patients have already lost 60-70% of DA-producing cells (Dauer & Przedborski 2003) proving that sporadic PD is diagnosed many years after the onset. It is therefore reasonable to expect that potential pharmacological treatments could be more effective if patients can benefit from it in the premotor phase.

Given the systemic nature of the disease, it is not surprising that many alterations of blood physiology have been described in PD patients (Kim et al., 2004; Shults and Haas, 2005; Bongioanni et al., 1996; Migliore et al., 2002; Petrozzi et al., 2002; Salman et al., 1999; Larumbe et al., 2001; Bessler et al., 1999). In this context, a blood test to predict PD would impact the ability to identify new treatments for this incurable disease. Furthermore, it could be applied to a large number of individuals since blood is commonly used in diagnostics for being easily accessible.

Gene expression analysis is a powerful tool to study complex diseases such PD and it has been extensively employed to find peripheral biomarkers (Papapetropoulos et al., 2007).

In the laboratory of Prof Gustincich, in collaboration with Dr Carninci at RIKEN, Yokohama, Japan, nanoCAGE technology has been previously used to find alterations in the blood transcriptome of 20 drug naïve *de novo* PD patients compared to 20 Healthy Controls (HC). NanoCAGE allows the identification of Transcription Start Sites (TSSs) and therefore of the associated promoters providing an unbias quantitative description of the cellular transcriptome targeting virtually any RNA molecule present in the sample.

The most up-regulated nanoCAGE tag in PD patients is located in the third intron of the gene Nitrogen Permease Regulator Like Protein 3 (Nprl3).

Nprl3 gene lies on the telomeric region of human chromosome 11 and contains in its

intron the major regulator elements of α globin (Hughes et al., 2005).

Neklesa and Davis in 2009 (in yeast) and Bar-Peled et al. in 2013 (in mammals) proved that Nprl3 is a component of a protein complex that inhibits mTORC1 activity.

In eukaryotes TOR is the major sensor of nutrients, energy and stress. Alterations in its pathway have been correlated with diseases and conditions where growth and homeostasis are compromised such as cancer, metabolic diseases and aging.

The aim of my PhD thesis was to identify the full-length transcript associated to the nanoCAGE tag, validate it, and to test whether it may represent a peripheral biomarker of PD.

Taking advantage of rapid amplification of cDNA ends (RACE) assay, I demonstrated that the tag represents an alternative Transcription Start Site of Nprl3 (TagNprl3). It is associated to a TCT motif (YC+1TYTYY) for initiation of transcription, which has been found to be specific for ribosomal protein coding genes and those involved in protein synthesis. The tag maps to a 29nt minisatellite that is found repeated 16 times in the reference genome. High-tag expression is associated to an allelic genomic variant of 13 repeats. To our knowledge this is the first time that a minisatellite variant is both a TSS and an expression quantitative trait locus (eQTL).

Unfortunately, high TagNprl3 expression resulted not to be correlated to PD but to heterozygosity. Furthermore, allelic frequencies were not correlated to PD.

I then showed that TagNprl3 is expressed in red blood cells (RBCs) both at mRNA and protein levels giving rise to an isoform truncated at the N-terminal. This is able to interact with its protein partner Nprl2 and its overexpression inhibits cell proliferation.

This work provides hints for Nprl3 protein function in blood and may suggest a testable hypothesis linking mTOR activity to genomic polymorphisms in modifier genes.

Introduction

Gene expression regulation

The fine regulation of gene expression in a living organism is extremely complex and requires several players for its correct tuning in time and space, which is fundamental for normal development and maintenance. Slight alterations could have dramatic cascade effects, from cellular level to whole organism, that can lead to loss of homeostasis and then to disease (Chen et al., 2008; Cookson et al., 2009; Emilsson et al., 2008).

Cells control and regulate their gene expression through different mechanisms, which can be summarized in six steps:

1) **Epigenetics**

   First defined by Conrad Waddington (1905–1975) as "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" (Goldberg et al., 2007) and comprehends all the heritable changes in gene expression that occur without DNA sequence alteration (Bird, 2007).

   The main epigenetic modifications are the chemical changes to cytosine residues of DNA (DNA methylation) and to histone proteins (Bernstein et al., 2007).

   Epigenetics control gene expression by changing chromatin structure: accessible chromatin is associated with actively transcribed gene and, viceversa, inaccessible chromatin is often associated with silent genes.

2) **Transcription**

   RNA polymerase is the main player in the transcription step that is modulated by several proteins called "activators" and "repressors" along with others transcription factors (TF). They can bind specific DNA sequences, namely enhancers and promoters, regulating gene expression through different mechanisms, for example stabilizing or blocking RNA polymerase, influencing histone protein modifications, recruiting other protein complexes. Furthermore, TF activity is in turn modulated via post-transcriptional modifications (PTMs) by specific enzymes in response to cellular stimuli.

3) **Post-translational modification (PTMs)**

PTMs convert a precursor messenger RNA (mRNA) into a mature mRNA and they are essential for a correct protein translation. They include splicing, 5'-capping and 3'-polyadenylation providing intron excision and RNA stabilization. RNA editing are rare events but there are evidences for which they can contribute to mRNA degradation (Agranat et al., 2008).

4) **RNA transport**

To regulate spatial gene expression, mRNAs are transported, localized and locally translated in several eukaryotic cell types (Kindler et al., 2005) and translation is repressed during transport to prevent ectopic expression (Kwon et al., 1999).

5) **Translation**

Translation is carried out by the ribosomes and is divided into three steps called initiation, elongation and termination. Two general modes of control can occur: global translation, in which the translation of most mRNAs in the cell is regulated, or mRNA-specific translation, in which a defined group of mRNA is affected (Gebauer and Hentze, 2004).

6) **mRNA degradation**

Eukaryotic mRNAs have a half-life that spans from few minutes to several days (Yu and Russell, 2001) and this strongly influences the rate of protein synthesis.

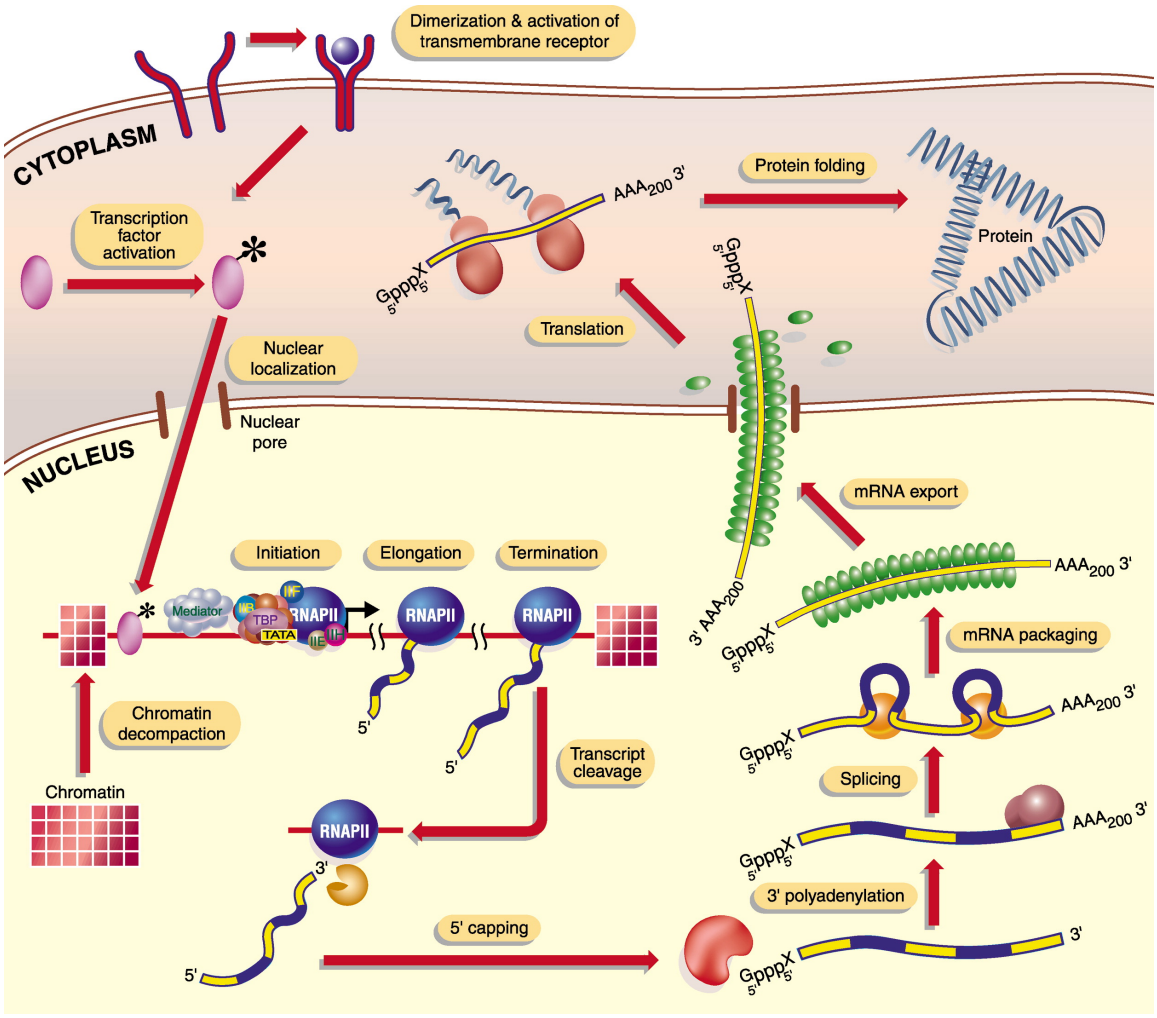

Figure 1. Different steps in the gene expression pathway (Orphanides and Reinberg, 2002).

Gene expression profiling

To identify the pattern of expression in a cell in a defined time and condition, gene expression profiling techniques have been developed.

Among them, DNA microarrays take advantage of DNA probes targeting annotated genes measuring their relative activity and, on the other hand, sequence-based techniques allow profiling of all active genes using a hypothesis-neutral approach.

However, despite the enormous potential of these technologies, the finding of an appropriate control in expression comparison, the heterogeneity of tissues and the statistical analysis of arrays represent challenging roadblocks to a faithful representation of biological activity. Finally, gene expression profiles must be also validated using complementary technologies as RT-PCRs (King and Sinha, 2001).

NanoCAGE

The use of nucleic acid sequencing has increased exponentially thanks to the new technologies that allow cheaper and faster results.

Next-generation sequencing (NGS) platforms can perform massively parallel sequencing delivering an entire genome in less than one day (Grada and Weinbrecht, 2013).

These technologies allow an unbiased description of the cellular transcriptome targeting virtually any RNA molecule present in the sample.

Tagging technologies, like SAGE and nanoCAGE, aim at the description of new transcripts types and their transcription start sites (TSSs).

Cap-Analysis of Gene Expression (CAGE) is a tagging technique developed at the RIKEN Institute in Japan. It is based on the production of short tag sequences close to the 5'-end of the transcript and the ligation of tags into groups of concatamers, followed by cloning and sequencing of ligation products (Gustincich et al., 2006; Kodzius et al., 2006; Shiraki et al., 2003). Through statistical calculations, assessing the number of tags related to a specific transcripts normalized per million of sequenced tags (tags per million, TPM), the information derived by CAGE data are quantitative.

The advantages of CAGE over other methodologies such as microarrays in identifying new promoters and new TSSs are becoming more evident as the sequencing technologies become more affordable. A major limitation of CAGE was the need for a large amount of starting RNAs. Recent advancements in 5'-end tagging and sequencing have made possible the development of a new technology of the CAGE type for total RNA quantities as low as 50 nanograms. This new technique, named "nanoCAGE", has been developed in the laboratories of S. Gustincich (SISSA) and P. Carninci (RIKEN). It is very powerful in the analyses of transcriptomes and promotomes of small RNA samples, for example those purified with Laser Capture Microdissection, Fluorescence-Activated Cell Sorters (FACS) and similar technologies.

TSS complexity

The complexity of transcription initiation has been historically difficult to grasp for the limits of the available technologies in pre-genomics era.

Differential TSS usage gives rise to a diverse repertory of 5'UTRs that may regulate mRNA translation, localization and stabilization.

A well-known example is the neurotrophin brain-derived neurotrophic factor (BDNF) gene that is characterized by eleven different 5'UTRs, generated by nine exons with different TSSs. BDNF is essential for neuronal survival, differentiation, and plasticity and is widely expressed in the brain. The laboratory of Prof Tongiorgi demonstrated that different TSS usage determines different subcellular localization of BDNF mRNAs that can be directed to distal dendrites or to soma and proximal dendrites (Chiaruttini et al., 2008; Pattabiraman et al., 2005; Tongiorgi et al., 2004).

For many years the study of differential TSS usage was carried out by analysing one gene at a time.

The purpose of CAGE developers was to create a technology to map TSSs and promoters at a genome-wide levels and that can be exploited to monitor transcriptional control in different tissues, cells and conditions. Deep-sequencing is necessary to detect all the active promoters in a given sample; simultaneously, the study of alternative promoters allow us to understand how a gene responds to distinct regulatory inputs, how mRNAs and protein isoforms are expressed in time and space and finally how gene isoforms that are preferentially expressed in a given tissue can be selectively targeted (Valen et al., 2009).

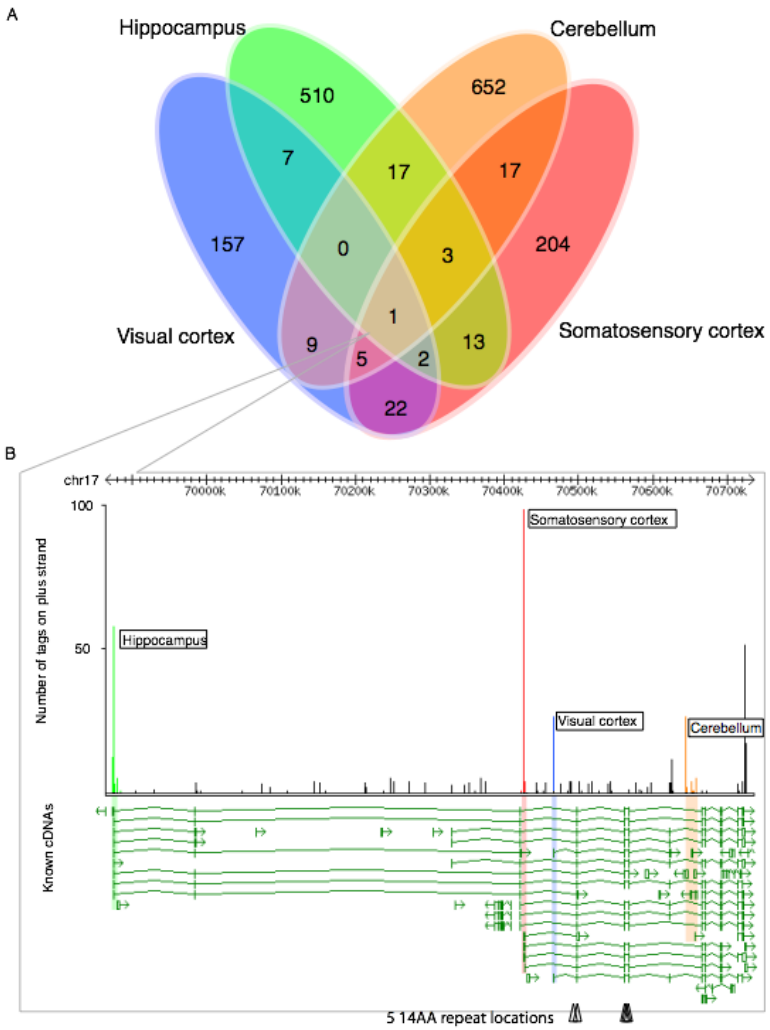

Figure 2. Example of differential TSSs usage in different brain tissues. A) The Venn diagram reports the number of genes that have at least one preferentially expressed promoter (PEP) in the four brain tissues or any combination PEPs of the four tissues. B) The Dlgap1 gene has four PEPs, one from each brain tissue (Valen et al., 2009).

Expression Quantitative Trait Loci (eQTL)

The development of techniques to study expression profiles has led to a better understanding of the biological processes that are implicated in human complex diseases. However, these approaches lead to analyses that are limited to the genome output (the transcribed RNA), omitting the genome variation that could be responsible for expression variations.

A genome-wide association study (GWAS) is an analysis on the genetic variants of a population to find variants associated to traits.

The aim of GWASs is usually to find traits associated to human diseases by analysing differences in Single Nucleotide Polymorphisms (SNPs) that represent about 90% of sequence variants in humans (Collins et al., 1998).

However, the risk of false-positive results, insufficient sample size, the biases due to case and control selection, the insensitivity to rare and structural variants and the lack of information on gene function are common problems (Pearson and Manolio, 2008).
To overcome the limitations of the two methodologies, in 2001 Jansen and Nap showed that by correlating genetic polymorphisms with expression profiles (RNA, protein or metabolites) it is possible to identify quantitative trait loci (QTL), introducing the concept of "genetical genomics".
In particular, mRNA levels are easily quantified by using microarrays and next generation sequencing (NGS) allowing the mapping of expression QTL (eQTL) that helped the dissection of the genetic basis of gene expression.
Genome-wide eQTL mapping has been conducted in many species (Brem et al., 2002; Rockman and Kruglyak, 2006; Schadt et al., 2008) and indeed is a powerful mean to study human complex diseases (Emilsson et al., 2008; Huang et al., 2007; Schadt et al., 2008; Zhong et al., 2010) that are caused by many genes, involving various biological pathways and varying in severity of symptoms and age of onset (Tabor et al., 2002).
eQTLs have been subdivided into *cis*-eQTLs and *trans*-eQTLs: in the former the genetic variant is located near the affected gene, in the latter is distant or on a different chromosome. *Cis*-eQTLs have usually higher effects on gene expression and are found close to the transcription start site (TSS) or in the genic sequence, probably altering transcription factors (TFs) binding or other *cis* regulatory elements. On the other hand, *trans*-eQTLs have usually smaller effect size and therefore they are difficult to detect (Westra and Franke, 2014).

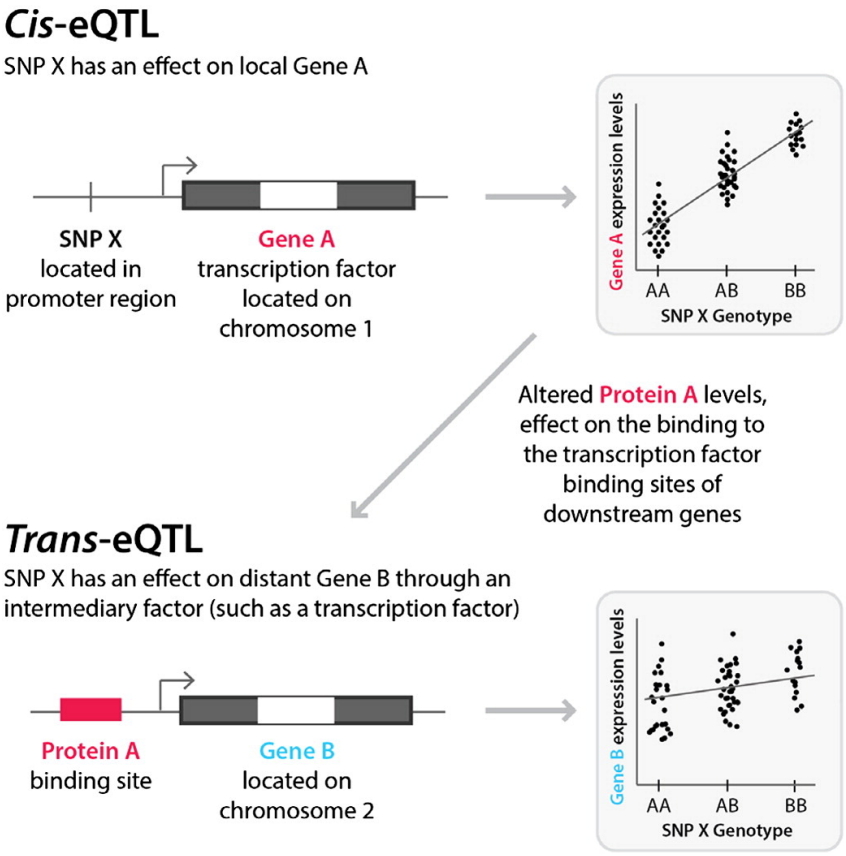

Figure 3. eQTLs can be either local effects (*cis*-eQTL), or distant, indirect effects (*trans*-eQTL) (Westra and Franke, 2014).

Repetitive elements in the mammalian genome

Despite being considered as "junk DNA" for many years, repetitive elements (RE) in the mammalian genome play an important functional role in genome organization and stability.

RE represent 50-70% of human genome and are simply defined as DNA sequences repeated multiple times (Padeken et al., 2015). They are classified into two groups, based on their origin, as follows:

1) **Tandem repeats**

   Tandem repeats are thought to origin by improper DNA replication and are highly variable among individuals.

   They are defined as pattern of DNA repeated one adjacent to another in a head to tail fashion. They can be classified based on the size of the repeated block. Microsatellites have a 2-5 bp core unit spanning up to hundreds of base pairs and are usually found dispersed throughout the genome. Minisatellites, also called Variable Number of Tandem Repeats (VNTR), have a unit length of 10-60 bp

with a conserved core sequence of 10-15 bp, spanning 0.1-15 kb and are usually found on telomeric regions. Satellite (centromeric and telomeric) tandem repeats have usually a 5-171 bp core unit and they can stretch to more than 100 kb (Padeken et al., 2015; Strachan and Read, 2004).

2) **Transposable elements**

As suggested by their name, transposable elements can mobilize and change their position in the genome and are subdivided into DNA transposons and RNA transposons depending on the intermediate form that is used for transposition.

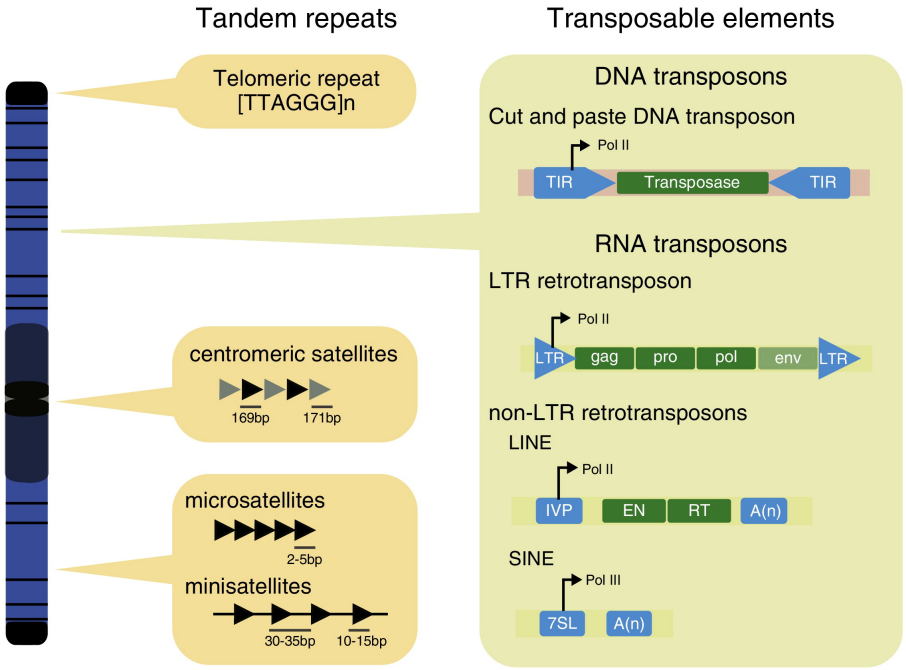

Figure 4. Schematic representation of RE classes, their distribution and their structural hallmarks (adapted from Padeken et al., 2015).

Minisatellites and human diseases

Wyman and White discovered the first minisatellite sequence in 1980. The majority of them are GC rich with a strong strand asymmetry exhibiting hypervariability and a high degree of polymorphism. However, most minisatellites are quite stable except few loci, called hypermutable minisatellites (Denoeud et al., 2003).

For this reason, they revealed to be useful as DNA fingerprinting in forensic applications, as first proposed by Jeffreys et al. in 1985. Nonetheless, the very nature of minisatellites makes them unsuitable for genotyping on array platforms, as it was possible for SNPs, because of the high variable number of repeats (Brookes, 2013).

For this reason, a genome-wide approach to study minisatellite variants correlated to human diseases was not possible although many studies have shown their important biological functions, such as regulating the level of expression of the nearby genes. Based on their position, minisatellite polymorphisms are subdivided into three classes, here described.

1) **Coding polymorphisms**

   Minisatellites polymorphisms occurring in the coding region of a gene are the best candidates for a functional activity, since they are translated into protein. One example is the minisatellite present in the third exon of the Dopamine Receptor D4 (DRD4) gene that has been associated with Attention Deficit Hyperactive Disorder (ADHD), response to clozapine in schizophrenia treatment (Shaikh et al., 1993) and other neurological disorders (Figure 5). The 48 bp minisatellite ranges from 2 to 11 repeats and, once translated, each tandem repeat adds 16 aminoacids to the third cytoplasmic loop of the protein (Chio et al., 1994; Van Tol et al., 1991). Albeit the function of the receptor is only partially influenced by the number of repeats and only in certain conditions, different DRD4 variants display differential sensitivity to dopamine chaperone effect in the cell (Van Craenenbroeck et al., 2005).

2) **Promoter polymorphisms**

   Transcriptional machinery binding to promoter is strongly influenced by its sequence and therefore minisatellites could have an important impact on the level of expression. Many polymorphisms have been found in promoter regions and some examples are the Serotonin Transporter (5HTT/SLC6A4) gene, the Insulin (INS) gene and the Monoamine Oxidase A (MAOA) gene.

   The 20-23 bp of 5HTT promoter are repeated 14 or 16 times and the short variant of the polymorphism reduces the transcriptional efficiency and can be partially related to anxiety-related traits (Lesch et al., 1996) which are related to increased risk of depression (Lotrich and Pollock, 2004). Moreover Pezawas and colleagues in 2005 showed with morphometric analysis that short-allele carriers have reduced gray matter volume and connectivity in limbic regions critical for processing of negative emotions, particularly perigenual cingulate and amygdala (Pezawas et al., 2005).

The 14-15 bp tandem repeat of INS promoter is present in variable number and in particular three classes of alleles have been defined based on its size: class I consists of 28-44 repeats, class II 45-137 and class III 138-159.

Class I alleles predispose in a recessive way to Type 1 Diabetes (T1D), while class III alleles are dominantly protective (Bennett et al., 1997).

A possible mechanism for this protective effect is the tolerance induction to insulin epitopes achieved by higher levels of insulin in the thymus promoted by class III alleles (Durinovic-Belló et al., 2005). On the contrary, class I alleles might predispose individuals to type 1 diabetes by lowering insulin expression in the thymus and rendering less tolerance induction (Cai et al., 2011).

MAOA gene has a 30 bp repeat sequence in the promoter region that is present 2, 3, 3.5, 4, 5 times and the variants are respectively called 2R, 3R, 3.5R, 4R, 5R.

Alleles 3.5R and 4R are transcribed 2-10 times more efficiently than 3R (Sabol et al., 1998). MAOA minisatellite has been associated with several behavioural disorders, including bipolar disorder, impulsivity and antisocial behaviour and in females the longer alleles with panic disorder (Brookes, 2013).

3) Other polymorphisms

Minisatellite polymorphisms can finally be present in intronic and untranslated regions of genes influencing RNA splicing, localization, stability and translational efficiency. An example is the Dopamine Transporter gene (DAT1/SLC6A3), which has a 3'untranslated region (UTR) minisatellite of 40 bp repeated 7-11 times.

The minisatellite allele can alter the level of expression of DAT1 although discordant data have been reported (Brookes, 2013).

The 10-repeat variant is abnormally efficient at the re-uptake process and underactive in the dopaminergic mesocorticolimbic and nigrostriatal pathways, suggesting reduced dopamine in mesolimbic and striatal pathways that are commonly implicated in major depressive disorder (MDD) and in ADHD (Gatt et al., 2015).

Nevertheless, the function of minisatellite polymorphisms present in non-coding regions is usually elusive and the alteration of nearby genes could be determined by the number of repeats and by the proteins that can target them. An example is the alpha-thalassaemia/mental retardation syndrome X-linked (ATRX) protein that alters gene expression by binding to G-rich tandem repeats. Law et al. in

2010 found that, in erythroid cells, ATRX mostly localizes on ψζ VNTR repeat in the alpha globin gene locus and, when it is mutated, induces down-regulation of the nearby genes explaining why patients exhibit alpha-thalassemia (Law et al., 2010).

Moreover, they demonstrated that patients with the same ATRX mutation could exhibit different severities of the disease because the inhibition is proportional to the repeat expansion.

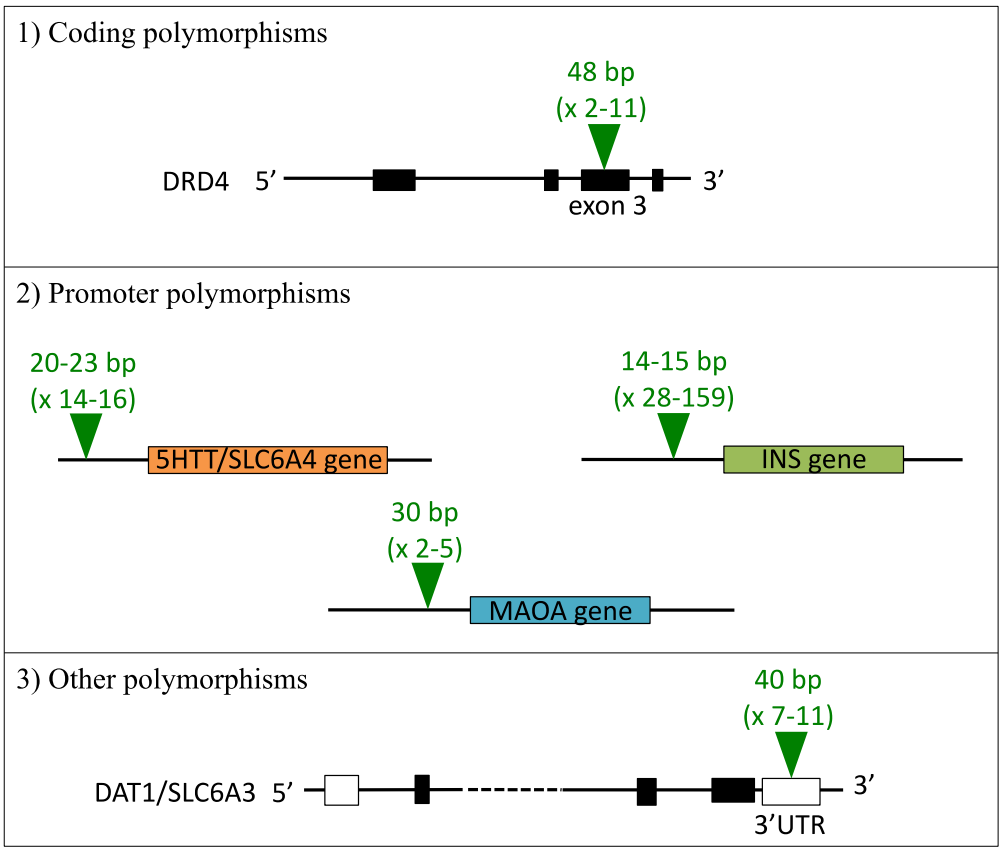

Figure 5. The three classes of minisatellite polymorphisms, based on their position relative to the influenced gene.

Parkinson's disease

Parkinson's disease (PD) is a common, complex, progressive neurodegenerative disorder first described in 1817 by doctor James Parkinson on his "An Essay on the Shaking Palsy" (Parkinson, 1817). The characterization of the pathology is still ongoing given its heterogeneity and complexity. While it was considered an exclusive motor disease for many years, recently numerous non-motor symptoms have been associated to PD and they can precede the motor onset by more than a decade, encouraging the seek of precocious markers of the pathology. In fact, an

early and better diagnosis is fundamental for PD diagnosis, giving the possibility to discover effective pharmacological treatments at the first stages of the disease in perspective of slowing its progression and providing a better and longer life for patients.

While its etiology is still unknown, PD is believed to develop from a complicated interplay between genetics and environment.

Importantly, PD remains with no cure and the existing therapies act only to alleviate symptoms.

PD clinical aspects

During the 1950s and 1960s, Arvid Carlsson (Nobel Prize in 2000) demonstrated that PD patients showed a massive loss of Dopamine (DA) in the brain, which can be clinically reverted by treatment with levodopa.

Nowadays PD is the second most common progressive neurodegenerative disorder (six million patients worldwide) after Alzheimer's disease (AD). It affects 1-2 % of all individuals above the age of 65 years old, increasing to 4-5% by the age of 85.

PD is characterized by six cardinal features: tremor at rest, rigidity, bradykinesia, hypokinesia and akinesia, flexed posture of neck, trunk and limbs, loss of postural reflexes and freezing phenomenon. The early symptoms of PD are usually alleviated by the treatment with levodopa and DA agonists. As PD advances from year to year, late symptoms such as flexed posture, loss of postural reflexes and freezing phenomenon as well as bradykinesia become resistant to pharmacological treatments. While motor symptoms dominate PD clinical features, many patients show non-motor symptoms. These include fatigue, depression, anxiety, sleep disturbances, constipation, bladder and other autonomic disturbances (sexual and gastrointestinal), sensory complaints, decreased motivation, apathy and a decline in cognition that can progress to dementia (Braak et al., 2004).

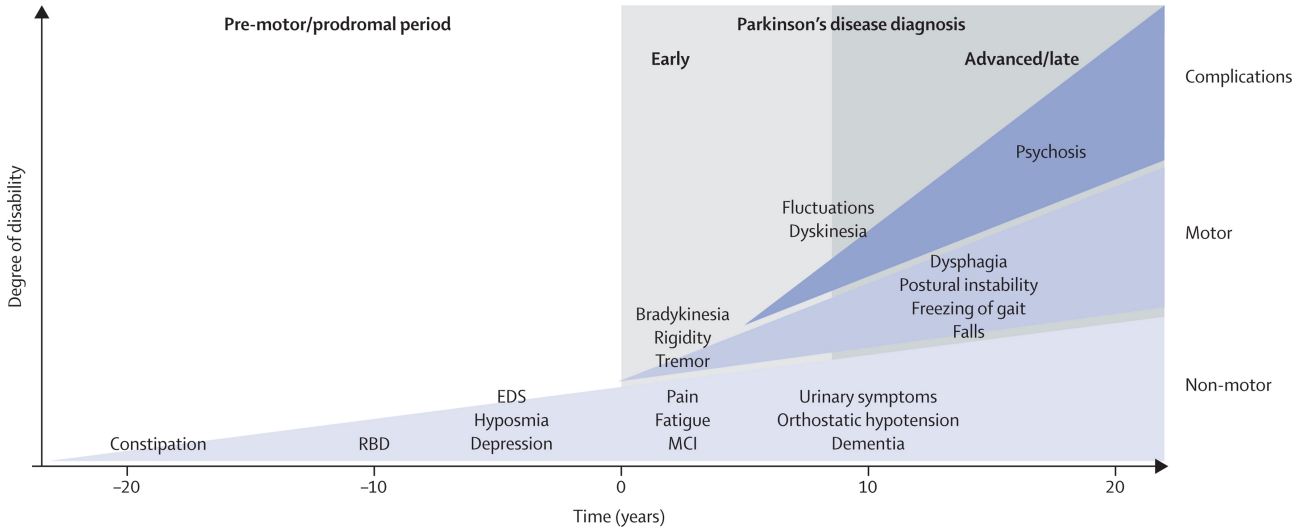

Figure 6. PD progression and relative clinical symptoms (Kalia and Lang, 2015).

Neuropathological features of PD

Analysis of *post mortem* PD brains show three major pathological features: depigmentation of the *substantia nigra* (SN) caused by the loss of DA neurons projecting in the striatum (principal cause of PD motor symptoms); presence of intraneuronal inclusions known as Lewy bodies within the surviving neurons of the SN as well as other brain regions and iron deposits.

The following features are considered PD hallmarks:

1) **Loss of midbrain DA neurons and consequent dopamine depletion in the striatum**

   Motor dysfunctions are key features of PD. Voluntary movements originate at motor cortex level: signals are sent through the encephalic trunk (*mesencephalon, pons and medulla*) to the brain stem and spinal cord. These signalling pathways are controlled by different sub-cortical structures (*thalamus, putamen and subthalamic nuclei*) that modulate movements thanks to a complex network of excitatory and inhibitory signals. These nuclei are, in turn, innervated and modulated by ventral midbrain DA neurons, subdivided into three main groups: A8 (retrorubral field; RRF), A9 (*substantia nigra pars compacta* SNc), and A10 (ventral tegmental area; VTA). The sensorimotor striatum involved in control of movement (*putamen*) is mainly innervated by DA cells of the ventral SNc whereas the limbic ventral striatum and the *thalamus* are targeted preferentially by VTA and dorsal SNc neurons. The characteristic motor symptoms of PD are due to the alteration of this movement control system, in particular, to the

selective, slow and progressive loss of DA neurons in the SNc. Their loss leads to a profound reduction in striatal dopamine that provokes an unbalance between excitatory and inhibitory trasmission in the basal ganglia-thalamocortical "motor" circuit (Dauer and Przedborski, 2003). The decreased striatal dopamine causes increased inhibitory output from the *globus pallidus* internal segment and *substantia nigra pars reticulata* (GPi/SNr) that, in turn, suppresses movement. On one hand, decreased striatal dopamine stimulation causes the direct reduction of the inhibition of the GPi/SNr; on the other hand, decreased dopamine inhibition causes increased inhibition of the *globus pallidus* external segment (GPe), resulting in blocking the inhibition of the subtalamic nucleus (STN). Increased STN output increases GPi/SNr inhibitory output to the thalamus that cannot regulate the activity of the motor cortex any longer.

The pattern of progressive cell loss is not homogeneous, but rather displays a complex topographical and regional organization: the degeneration starts from the most lateral part of SNc and spreads to the most medialis region. Furthermore, nigrostriatal projections to the putamen are more sensitive than those to caudate and nucleus accumbens regions, whereas VTA projections to the ventral striatum are selectively spared in PD patients.

While loss of the nigral DA neurons and terminals are responsible for the movement disorders associated to PD, it has become clear that additional neuronal populations throughout the brain are also affected in the disease (Braak et al., 2004).

2) Lewy bodies

Typical pathological hallmarks of the disease are Lewy Bodies (LBs), discovered in 1919 by Trétiakoff, and dystrophic neuritis, also called Lewy Neurites (LNs) first described by Friederich Lewy in the dorsal motor nucleus of the vagus nerve and in the *substantia innominata* from patients with PD.

LBs and LNs are present in PD patients' brains in some of the surviving DA neurons and in other regions like dorsal motor nucleus of the vagus, locus ceruleus, thalamus, amygdala, olfactory nuclei, pediculopontine nucleus, and cerebral cortex. In late stages of PD, when patients manifest all motor and cognitive clinical symptoms, inclusion bodies decorate the entire neocortex comprising the sensory association area (Braak et al., 2004). LBs are round cytoplasmatic eosinophilic inclusions, 5–25 μm in diameter. The ultrastructure

show that they are composed of a dense core of filamentous and granular material, surrounded by radially orientated filaments of 10–20 nm in diameter.

LBs are composed by the accumulation of cytoplasmic aggregates containing a variety of proteins, like α-synuclein, the major component, parkin and UCHL1 (genes whose mutations have been correlated with PD) and many others like Hsp70, ubiquitin and components of the proteasome (Spillantini and Goedert, 2013).

LNs are abnormal neurites that contain the same proteins and abnormal filaments found in LBs. The mechanism causing the abnormal accumulation of proteins in LBs and LNs is not yet known and their toxic or protective role in neurodegenerative process is still matter of debate because until now there is no clear correlation between inclusion formation and neuronal cell death. If LBs are considered toxic, cytoplasmic protein aggregates may interfere with intracellular trafficking or sequestrate proteins important for cell survival. On the other hand, LBs are present in surviving cells that seem to be healthier, by morphological and biochemical analysis, than neighbouring cells (Tompkins et al., 1997). Moreover, LBs have been also found in people without evident neuronal loss or clinical signs of PD, so they could be a protective structure in which proteins accumulate after the pathological failure of protein degradation.

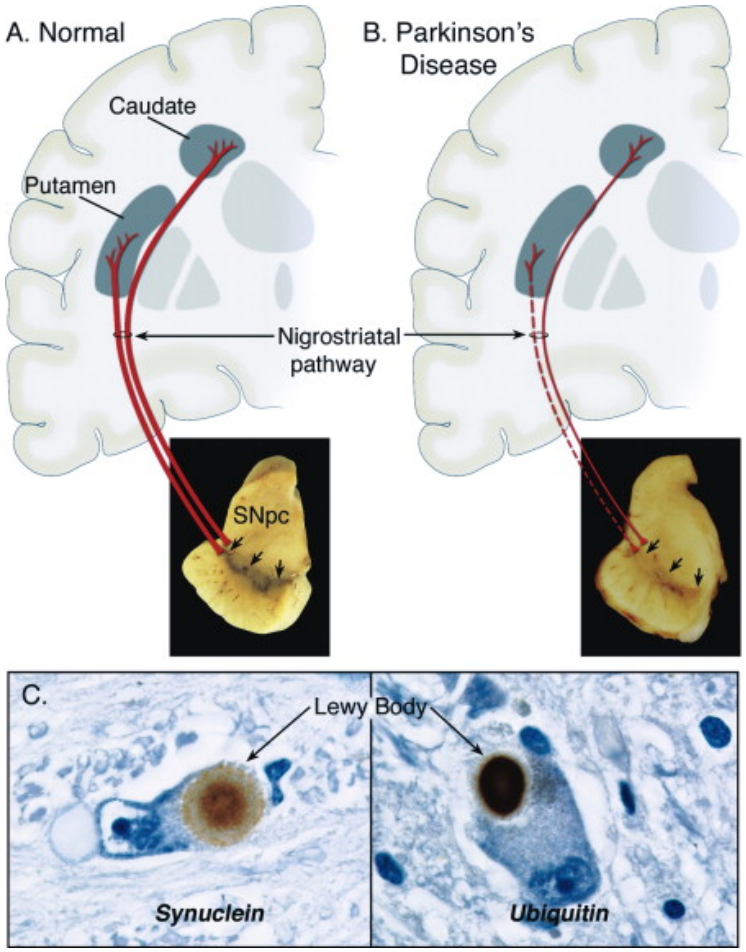

Figure 7. A) Schematic representation of the normal nigrostriatal pathway (in red). SNc neurons project (thick solid red lines) to the basal ganglia and synapse in the striatum. B) In PD, the nigrostriatal pathway degenerates: there is a marked loss of DA neurons that project to the putamen (dashed line) and a much more modest loss of those that project to the caudate (thin red solid line) C) Immunohistochemical staining of Lewy Bodies found in DA neurons of PD patients in *substantia nigra pars compacta.* Unsoluble fibrous component of aggregates are detected with anti α-synuclein antibody and anti-ubiquitin (Dauer and Przedborski, 2003).

3) Iron accumulation

Iron is important for many biological processes. It is coordinated by heme, and it is an essential component of cytochromes as well as a cofactor of many different enzymes involved in the normal function of neuronal tissue. For example, tyrosine hydroxylase, which is required for dopamine synthesis, is a non-heme iron enzyme.

Large amounts of iron are sequestered in neuromelanin granules in the DA neurons of the *substantia nigra* and the noradrenergic neurons of the locus ceruleus. Neuromelanin is synthesized by the oxidation of excess cytosolic catecholamines that are not accumulated in synaptic vesicles by vesicular monoamine transporter-2 (VMAT2). Interestingly, it binds iron avidly.

Iron accumulation in *substantia nigra* and *globus pallidus* has been also described as pathological hallmark of PD. The role of iron in neurodegenerative diseases, particularly PD, has not been clarified yet but different evidences show how its role, in combination with aging, could be critical for neurodegeneration of DA cells. To underline the toxic effect of iron in PD pathology it is important to mention that iron chelators, when administered prior to the exogenous toxin mimicking PD pathogenesis, appear neuroprotective in mice and non-human primates (Kaur et al., 2003).

PD etiology

PD is a multifactorial disease caused by both environmental and genetic factors. The cause of sporadic PD is unknown, but the environmental hypothesis was dominant for much of the 20th century for the discovery that an exogenous toxin can mimic the clinical and pathological features of PD. In 1983 Langston et al. discovered that people intoxicated with 1-methyl-4-phenyl-1,2,3,6-tetrahydropyrydine (MPTP) developed a syndrome nearly identical to PD (Langston et al., 1983). Moreover, human epidemiological studies have correlated residence in rural environment and related exposure to herbicide (i.e. Paraquat) (Tanner, 1989) and pesticides (i.e. Rotenone) (Betarbet et al., 2000) with an elevated risk for PD.

Although PD was long considered a non-genetic disorder of sporadic origin, 5–10% of patients are now known to have monogenic forms of the disease.

The discovery of the genes involved and of their pathological mutations led to a better understanding of the mechanisms underlying this complex disorder.

Clinical genetic testing can be helpful only for a limited number of cases with typical PD. However in atypical parkinsonian syndromes it could reveal mutations that are usually associated with other inherited diseases such as the spinocerebellar ataxias or the frontotemporal dementia (Gasser, 2015).

	Gene	Chromosomal position	Clinical characteristics
Autosomal-dominant PD			
PARK1	*SNCA*	4q21	• Different point mutations • Parkinsonism and dementia • Onset early (A53T) to late (A30P) depending on mutation
PARK4	*SNCA*	4q21	• duplication or triplication of the wild-type gene • parkinsonism and dementia • onset early (triplication) to late (duplication)
PARK8	*LRRK2*	12q12	• point mutations • similar to sporadic PD, mostly late onset
PARK17	*VPS35*	16q11	• similar to sporadic PD
Autosomal-recessive PD			
PARK2	*Parkin*	6q25	• nonsense and missense mutations, deletions, duplications • dystonia and dyskinesias common, slow progression
PARK6	*PINK1*	1p35	• similar to *parkin*-associated PD
PARK7	*DJ-1*	1p36	• similar to *parkin*-associated PD
Complex syndromes with parkinsonism			
PARK9	*ATP13A2*	1p36	• Parkinsonism, pyramidal syndrome, dementia
PARK14	*PLA2G6*	22q13.1	• dystonia-parkinsonism, pyramidal syndrome, dementia
PARK15	*FBXO7*	22q12	• parkinsonism, pyramidal syndrome, dementia
n.a.	*DNAJC6*	1p31.3	• parkinsonism, pyramidal syndrome, cognitive impairment, seizures
n.a.	*SYNJ1*	21q22.11	• parkinsonism, cognitive impairment, seizures

Table 1. Monogenic forms of PD (Gasser, 2015).

PD pathogenesis

Whatever the initial insult is, the analysis of human *post mortem* brains of PD patients as well as studies on PD animal models (neurotoxin-treated or genetic-modified) suggest the involvement of two major, possibly interconnected, pathways in dopaminergic neurodegeneration:

1) **Protein misfolding and aggregation**

   Alteration of protein folding, ubiquitin proteasome system (UPS) and autophagy are considered among the main molecular mechanisms in aggregate formation. The most important proteins involved in the correct folding of polypeptides are chaperons: this highly conserved class of proteins prevents inappropriate interactions within and between non-native polypeptides, enhances the efficiency of *de novo* protein folding and promote the refolding of stress-induced misfolded proteins (Hartl and Hayer-Hartl, 2002). When misfolded proteins cannot be "repaired" by chaperons, they have to be degraded. In the classical mechanism of degradation, proteins are poly-ubiquitinated to be degraded by the proteasome. Ubiquitination of a substrate requires a cascade of enzymes: ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzymes (E2s) and ubiquitin-protein ligases (E3s) that are responsible for highly specific target recognition in this system through physical interactions with the substrate.

A relatively new discovered mechanism through which proteins can be degraded is autophagy. Most long-lived proteins, cytoplasmic constituents, including organelles, are sequestered into double-membrane autophagosomes, which subsequently fuse with lysosomes where their contents are degraded thanks to proteolytic enzymes.

2) **Mitochondrial dysfunctions: oxidative stress and alterated fusion/fission machinery**

A consequence of the mitochondrial dysfunction is oxidative stress and *vice versa*. Ninety-five percent of the molecular oxygen is metabolized within the mitochondria by the electron-transport chain: thus, mitochondria are highly exposed and damaged by oxidative stress. This leads to a more intense and perpetuating cycle in which reactive oxygen species (ROS) are generated.

ROS cause functional alterations in proteins, lipids and DNA. Lipid damage leads to loss of membrane integrity and ions permeability, promoting excitotoxicity (Halliwell, 1992). Although ROS levels cannot be directly measured, the assessment of their reaction products and of the resulting damage in *post mortem* tissues is as an indirect index of their levels (Foley and Riederer, 2000).

Over the last several decades, evidence has accumulated that mitochondrial dysfunction is associated with PD. A mild deficiency in mitochondrial respiratory electron transport chain NADH dehydrogenase activity (Complex I) was first found in the *substantia nigra* of patients with PD (Mann et al., 1994; Schapira et al., 1989), followed by studies identifying a similar Complex I deficit in platelets (Blandini et al., 1998; Haas et al., 1995; Krige et al., 1992), lymphocytes (Barroso et al., 1993; Yoshino et al., 1992), and, less consistently, in muscle tissue (Penn et al., 1995; Taylor et al., 1994). The hypothesis that oxidative stress plays a role in the pathogenesis of PD was also previously proposed by the discovery that MPTP blocks the mitochondrial electron transport chain by inhibiting complex I (Nicklas et al., 1987). MPTP is highly lipophilic, and it crosses the blood-brain barrier within minutes. Once in the brain, it is oxidized to 1- methyl-4 phenyl-2,3-dihydropyridinium ($MPDP^+$) by monoamine oxidase B (MAO B) in glial cells and serotoninergic neurons and then is spontaneously oxidized to MPP^+.

Due to its high affinity for the DA transporter (DAT), it is selectively accumulated in DA neurons, where it causes toxicity and neuronal death through complex I inhibition. Similar toxic effects are produced by the common herbicide 1,1'-dimethyl-4,4'-5 bipyridinium (Paraquat) coupled with the administration of the fungicide manganese ethylenepistithiocarbamate (Maneb). While Paraquat, which is structurally similar to MPTP, blocks the mitochondrial complex I, Maneb inhibits the mitochondrial complex III. Rotenone, that freely crosses cellular membranes and accumulates in subcellular organelles such as mitochondria, impairs oxidative phosphorylation by inhibiting complex I of the electron transport chain and leads to neuropathologic and behavioural changes in rats similar to human PD (Alam and Schmidt, 2002). Other important and complex processes involved in mitochondrial dysfunction are alteration in mitochondrial fusion and fission. Mitochondrial fission and fusion are important mechanisms that maintain the integrity of mitochondria, their electrical and biochemical connectivity, their turnover, and the segregation, stabilization, and protection of mitochondrial DNA (mtDNA). When altered, they cause morphological and functional mitochondrial abnormalities. The correlation between oxidative stress and fission/fusion machinery is still matter of debate but it seems that the presence of ROS can alter mitochondria morphology due to a disregulation of genes involved in this process like Mfn1, Fis1 and Drp1 (Liot et al., 2009; Zhang et al., 2011). These mechanisms are of high interest in neurons because these cells have unique features such as post-mitotic state and processes with higher energy requirements. Therefore, a fine regulation of dynamic fission and fusion processes are particularly important. In addition, mitochondrial fission/fusion machinery is intimately and critically involved in the formation of synapses and dendritic spines; preventing mitochondrial fission leads to a loss of mitochondria from dendritic spines and a reduction of synapse formation, whereas increasing fission increases synapse formation (Li et al., 2004).

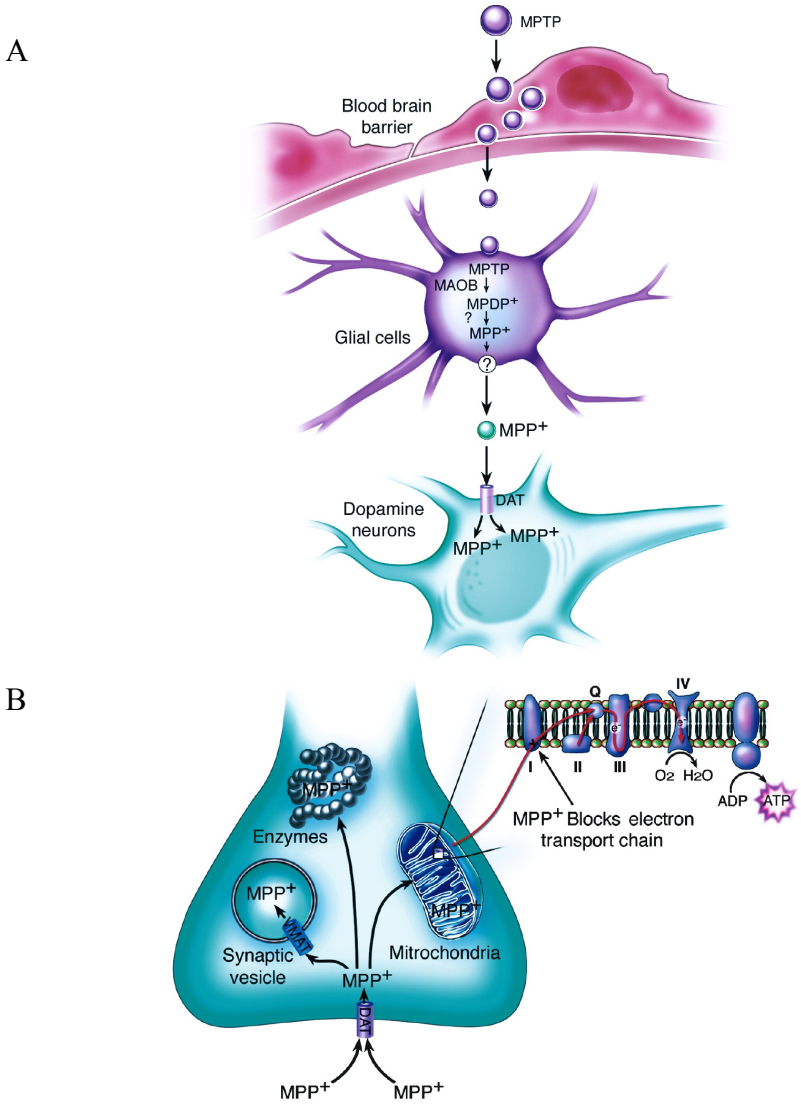

Figure 8. A) Schematic representation of the MPTP metabolism and B) of MPP+ intracellular pathway (Dauer and Przedborski, 2003).

In recent years the study of familial forms of PD proved the importance of mitochondria in the pathogenesis of the disease. Initial studies were concentrated on the high level of oxidative stress in the DA neurons of SN and on proteins such as DJ-1 (PARK7) contributing to the elimination of ROS (Clements et al., 2006; Taira et al., 2004). Recently, the analysis focused on the mechanisms of fusion and fission that, if altered, lead to oxidative stress. In particular, genes such as DJ-1, Parkin (PARK2) and PINK1 (PARK6) are involved in controlling mitochondrial morphology by regulating fusion and fission (Dodson and Guo, 2007). Finally, it is important to add that, although the molecular mechanisms are independent, the cellular pathways that control protein folding, protein

degradation and mitochondrial function are closely related and genes mutated in PD have been linked to several of these signalling pathways.

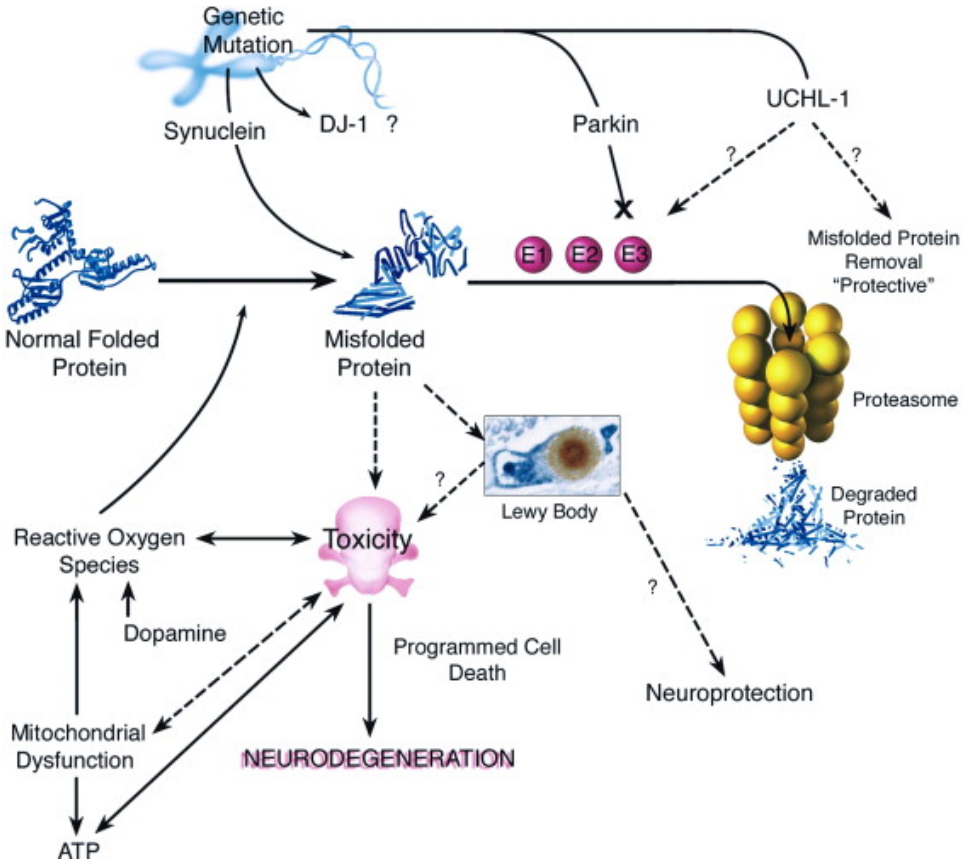

Figure 9. Mechanism of neurodegeneration: linking oxidative stress and mitochondrial damage to UPS impairment and aggregates formation (Dauer and Przedborski, 2003).

PD diagnosis

PD diagnosis, at the early stages of the disease, is nowadays particularly difficult because it is mainly based on motor symptoms and on patient's neurological history. The Society Brain Bank criteria are used in the clinic and they are subdivided in three steps here described. These criteria have an expected accuracy of up to 90% (Hughes et al., 2001).

- Step 1. Diagnosis of Parkinsoniam Syndrome: bradykinesia and at least one among muscular rigidity, 4-6 Hz rest tremor, postural instability.
- Step 2. Exclusion criteria for PD-like history of repeated strokes with stepwise progression of parkinsonian features, history of repeated head injury, history of definite encephalitis, oculogyric crises, neuroleptic treatment at onset of symptoms, more than one affected relative, sustained

remission, strictly unilateral features after 3 years, supranuclear gaze palsy, cerebellar signs, early severe autonomic involvement, early severe dementia with disturbances of memory, language, and praxis, Babinski sign, presence of cerebral tumor or communication hydrocephalus on imaging study, negative response to large doses of levodopa in absence of malabsorption, MPTP exposure.

- Step 3. Supportive prospective positive criteria for PD. Three or more required for diagnosis of definite PD in combination with step one: unilateral onset, rest tremor present, progressive disorder, persistent asymmetry affecting side of onset most, excellent response (70-100%) to levodopa, severe levodopa-induced chorea, levodopa response for 5 years or more, clinical course of ten years or more.

No standard tests are available nowadays to diagnose PD. Although the motor criteria together with the presence of neuronal loss in the SNc and of Lewy bodies in the surviving neurons are established as hallmarks of the pathology, the International Parkinson and Movement Disorder Society has pointed the needs of new criteria (Berg et al., 2013). By the time motor symptoms appear, PD patients have already lost 60-70% of the DA-producing cells in the SNc (Dauer and Przedborski, 2003) and therefore in most cases the disease is diagnosed many years after the onset. The pharmacological treatment can be more effective if patients can benefit from it in the premotor phase.

It is therefore imperative to identify peripheral biomarkers of PD. A biomarker is defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention" (from National Institutes of Health Biomarkers Definitions Working Group).

To this purpose, many efforts have been made to find a "PD signature" using different approaches that are summarized in Figure 10. Some biomarkers should predict if a patient is responding well to therapy, helping to monitor the disease progression in an objective way and providing a "personalized medicine" taking advantage of different approaches.

Clinical standard methods can measure olfactory impairment (for example with the University of Pennsylvania's smell identification test) and the rapid eye movement sleep behaviour disorder (with polysomnography).

Imaging technologies like positron emission tomography (PET) and single photon emission computed tomography (SPECT) could be helpful to measure SNc DA neuronal loss however they cannot distinguish PD from other diseases with SNc degeneration. Furthermore, PET and SPECT imaging are abnormal only when a significant loss has been reached.

Cardiac scintigraphy could be helpful to differentiate PD from multiple system atrophy (MSA), which often presents with some of the same symptoms, demonstrating that in PD with autonomic failure, but not in MSA, there is a myocardial postganglionic sympathetic dysfunction (Courbon et al., 2003).

Pathological approaches include colonic biopsy and skin biopsy both with α-synuclein staining that has been demonstrated to accumulate within the enteric peripheral nervous system (Visanji et al., 2014) and in the skin (Donadio et al., 2014) of PD patients.

Biochemical assays on biological fluids represent a relative non-invasive promising test although good candidates are still needed.

Finally, the genetic approach can be useful in patients with family members with a known monogenic form of PD.

A single methodology should not be exhaustive for the diagnosis and therefore a combination of these technologies might be required (Kalia and Lang, 2015).

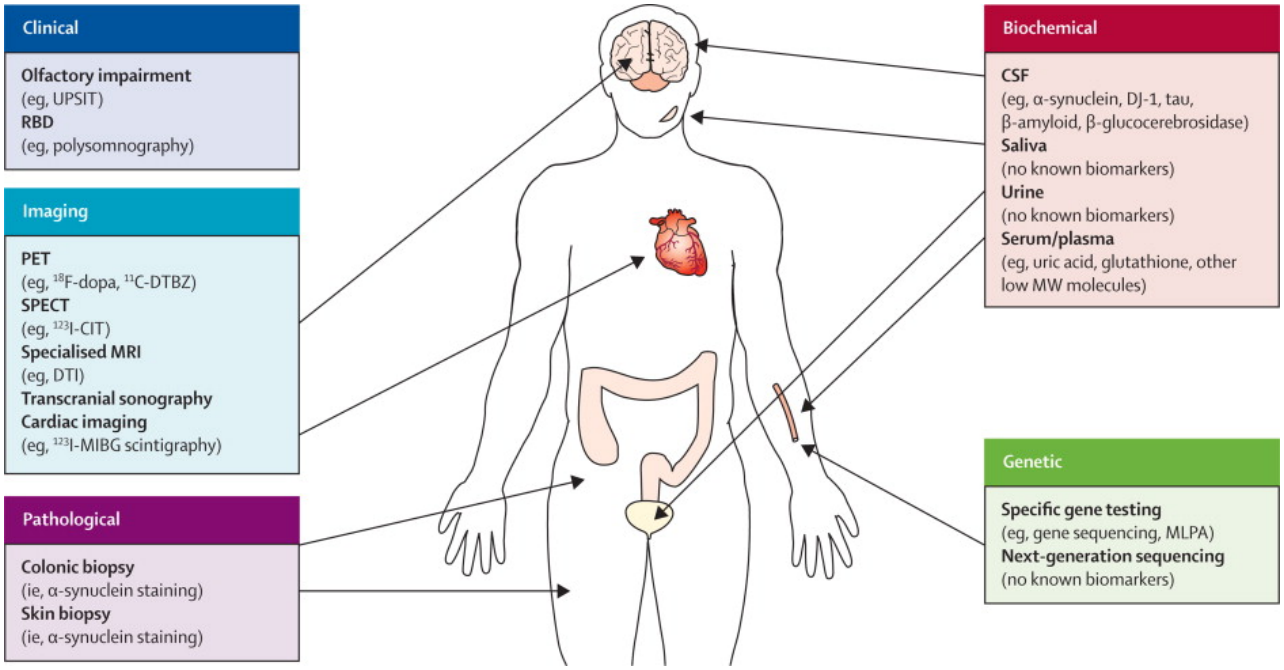

Figure 10. Potential biomarkers for diagnosis of PD. Many potential PD biomarkers are currently under investigation and they can be divided into clinical, imaging, pathological, biochemical and genetic. For an accurate diagnosis a combination of biomarkers is essential. DTBZ=dihydrotetrabenazine; CSF=cerebrospinal fluid; DTI=diffusion tensor imaging;

CIT=2β-carbomethoxy-3β-(4-iodophenyl)tropane; MIBG=metaiodobenzylguanidine; MLPA=multiplex ligation-dependent probe amplification; MW=molecular weight; PET=positron emission tomography; RBD=rapid eye movement sleep behaviour disorder; SPECT=single photon emission computed tomography; UPSIT=University of Pennsylvania's smell identification test (Kalia and Lang, 2015).

Blood Transcriptomics in PD

A blood test to diagnose PD would be ideal because blood is an easily accessible tissue.

Gene expression analysis is a powerful tool to study complex disorders such PD and it has been used to find biomarkers (Papapetropoulos et al., 2007). Given the systemic nature of the disease, it is not surprising that many works described alterations of blood physiology in PD patients, such as upregulation of α-synuclein at mRNA and protein levels (Kim et al., 2004), alteration in coenzyme Q10 (CoQ10) redox state (Shults and Haas, 2005), increased monoamine oxidase (MAO) molecular activity (Bongioanni et al., 1996), cytogenetic alterations in blood lymphocytes (Migliore et al., 2002; Petrozzi et al., 2002), decreased phagocytic function of blood polymorphonuclears (Salman et al., 1999), lowered levels of glutathione (Larumbe Ilundáin et al., 2001), higher levels of interleukin 1 beta (Bessler et al., 1999).

A short but significant list of recent microarray-based studies has used human blood as RNA source to look for differentially expressed genes in sporadic and genetic PD patients (Karlsson et al., 2013; Kedmi et al., 2011; Mutez et al., 2011, 2014; Potashkin et al., 2012; Scherzer et al., 2007; Soreq et al., 2008). In the pioneering work by Scherzer et al., 22 genes were identified with microarray profiling of whole blood of 50 PD patients, among which 9 *de novo* subjects. A risk marker given by 8 genes (VDR, HIP2, CLTB, FPRL2, CA12, CEACAM4, ACRV1, and UTX) predicted PD and was not biased by dopamine replacement therapy. The analysis of a genetically homogenous population of 88 Ashkenazi patients (Kedmi et al., 2011), including 20 *de novo*, evidenced for the first time the decreased expression of B cells-related genes in PD. Karlsson et al. analyzed samples from 79 PD subjects at different stages of the disease, including 23 *de novo* patients and relative controls, proposing a classifier predicting sporadic and *de novo* PD. LRPPRC, BCL2 and SRSF8 were shared with the 22 genes list as in Scherzer et al. Furthermore, it presented HSPA8 and UBE2K/HIP2 in common with Molochnikov et al., 2012.

Potashkin et al. took advantage of splice variant-specific microarrays to identify a biosignature composed of 13 mRNAs (c5orf4, wls, macf1, prg3, eftud2, pkm2, slc14a1-s, slc14a1-l, mpp1, copz1, znf160, map4k1 and znf134) whose expression is altered in peripheral blood of early-stage PD patients. Recently, they identified two novel longitudinally markers (HNF4 and PTB1) by means of network-based and transcriptomic meta-analyses (Santiago and Potashkin, 2015). Interestingly, gene expression analysis of peripheral blood mononuclear cells from 20 sporadic PD patients and 9 individuals, heterozygous for the LRRK2 G2019S mutation, showed deregulation of the immune system, endocytosis and eukaryotic initiation factor 2 signaling (Mutez et al., 2014).

Although there are a number of promising gene signatures, blood transcriptomics have not yet delivered the expected results for biomarker discovery in PD. One of the major concerns is the scarce overlap among candidate genes' lists of these studies. Variances in the procedures for collection, processing and analysis of samples may strongly limit the reproducibility of gene expression data. Importantly, these differences may also be explained in biological terms. First, genetic variations in human populations may lead to diversity in transcriptional changes in disease. Furthermore, the majority of these works analyzed peripheral blood samples from sporadic PD patients at different stages of the disease and under pharmacotherapy raising the questions of whether changes are related to the disease stage, therapy or both. Finally, it is now clear that PD is a systemic and a highly heterogeneous disease, as classified according to distinct clinical subtypes (Thenganatt and Jankovic, 2014).

On the other hand, blood transcriptomics studies identify a common repertory of enriched GO biological terms as altered in PD. These include "neuronal apoptosis", "mitochondrial dysfunction", "leukocyte activation" and "deregulation of the immune system".

Nitrogen permease regulator-like 3 (Nprl3)

Nitrogen Permease Regulator-Like 3 (Nprl3) gene was discovered in 1995 by the group of Higgs who described its position and sequence. It lies on the telomeric region of human chromosome 11 and contains in its intron the major regulator elements of α globin, which cluster is adjacent (Vyas et al., 1995).

The same group studied the sequence homologies by comparative analysis between human, mouse, chicken, and zebrafish revealing a syntenic region that includes the whole cluster of α-globin genes and the extended 5′-terminal region where Nprl3 gene is located (Flint et al., 2001; Tufarelli et al., 2004). It is transcribed in the direction opposite to the direction of the globin gene transcription and for many years was represented in different database under the name of C16orf35. Neklesa & Davis in 2009 changed its name in Nprl3 discovering the homology with Nitrogen Permease Regulator 3 (Npr3) yeast protein (Neklesa and Davis, 2009).

Vyas et al. 1995 reported the characterization of Nprl3 gene adjacent to the alpha globin cluster on chromosome 16p13.3, containing 15 exons and spanning about 55 kb. Its sequence and position has been conserved for at least 270 million years.

To date 5 different Nprl3 isoforms are annotated in the databases and their main differences are in the 5' region (5'UTR and position of the translational starting codon).

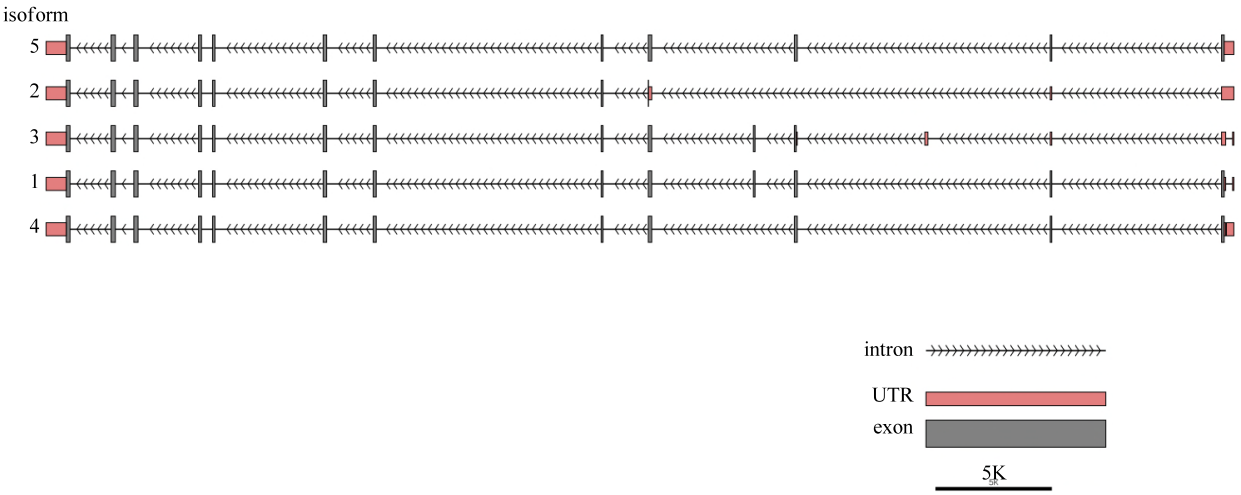

Figure 11. Schematic representation of Nprl3 gene isoforms. Transcription direction is from right to left since the gene is on the (-) strand of DNA.

Nprl3 expression regulation has been not yet extensively studied, however, the isoform 1 is considered as the canonical one. Therefore, in this thesis, when the isoform is not specified, "Nprl3" refers to isoform 1.

Nprl3 expression is ubiquitous and it has been found upregulated in human and mouse erythroid cells, where high level of α-globin expression occurs (Kowalczyk et al., 2012; Lower et al., 2009).

The mammalian alpha globin locus lies within a region of 135–155 kb of conserved

synteny, and human gene arrangement is 5′-ζ-αD-α2-α1-θ-3′, where α1 and α2 are hemoglobin alpha genes, ζ gene encoder for an alpha-like protein expressed during embryonic development, θ and αD are two pseudogenes (Figure 12) (Vernimmen, 2014).

Interestingly, Nprl3 gene is fundamental for hemoglobin alpha (HBA) genes expression: the evolutionarily conserved remote elements which regulate their expression correspond to erythroid-specific DNase1 hypersensitive sites located 40-60 kb upstream of the α genes (Hughes et al., 2005). These elements are called erythroid-specific multispecies conserved sequences (MCS) and they are numbered MCS-R1 to MCS-R4. Three of these elements (MCS-R1, MCS-R2, and MCS-R3) are located in Nprl3 introns and MCS-R4 lies upstream of the promoter of that gene (Vernimmen, 2014).

MCS-R2 (called also HS-40) lies in the fifth intron of Nprl3 gene and its deletion leads to almost complete down-regulation of alpha-gene expression causing severe thalassemia (Bernet et al., 1995).

Quantitative chromosome conformation capture analysis performed by Vernimmen and colleagues in 2007 demonstrated that a looped structure that bridges the remote regulatory elements to the promoter is involved in this distal regulation. This regulatory sequence does not influence expression of Nprl3 gene, which has a GC-rich promoter (unlike the TATA-type promoters of the alpha globin loci) (Vernimmen et al., 2007).

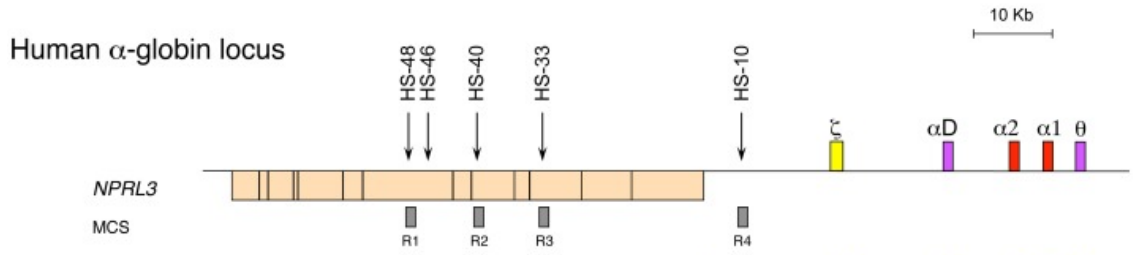

Figure 12. The chromosomal organization of human alpha-globin locus. Nprl3 gene lies upstream to the alpha-globin genes and pseudogenes (ζ, αD, α2, α1, θ) on the opposite strand of DNA. It contains the major regulators of alpha-globin expression that are indicated as R1-R4 multispecies conserved sequences (MCS) that correspond to DNA hypersensitive sites (HSs) (Vernimmen, 2014).

Nprl3 protein

Although Nprl3 gene discovery dates back to 1990s, for many years the deduced protein has not been studied even if it is highly conserved in metazoa, suggesting a conserved function.

Lunardi et al., 2009 demonstrated that Nprl3 encodes for a 64 kDa nucleo-cytoplasmic protein. By overexpressing it, they noticed that nuclear Nprl3 may be localized into the PML nuclear bodies, the matrix-associated domains that are proposed to recruit many proteins and regulate different nuclear functions such as DNA replication, transcription or epigenetic silencing (Stuurman et al., 1990).

They showed that Nprl3 overexpression inhibits cell proliferation in a human osteosarcoma cell line and that Nprl3 protein could bind all p53-family proteins, with a higher affinity for p73. Indeed the interaction with these transcription factors may suggest a role of Nprl3 in cell cycle control, tumor suppression and development.

When Neklesa & Davis screened the yeast genome, searching for genes responsible of amino acid sensing and starvation, they discovered that the evolutionary conserved Npr2/Npr3 complex is responsible of TORC1 activity in response to amino acid deprivation. Furthermore, they suggest their function is conserved since they proved Nprl2/Nprl3 interaction in mammals (Neklesa and Davis, 2009).

In 2013, Levine et al. identified two functional domains in Nprl2 and Nprl3 proteins: a longin domain (LD) at the N-terminal and two or more helix turn helix (HTH) domains at the C-terminal. The latter might bind DNA, RNA or protein partners. The first one comprises a five-stranded β-sheet core sandwiched between an α-helix on one side and two α-helices on the other. LDs have been found in many GDP/GTP exchange factors (GEF) that activate Rab-GTPases (Levine et al., 2013). The authors suggest that LDs could act as a platform for a GTPase involved in TORC1 signalling.

Notably, studying the genetic determinants of haemolysis in sickle cell anemia Milton et al. found an association between a SNP in NPRL3 first intron (rs7203560) and haemolytic score. They showed that SNPs close to MCS are in strong linkage disequilibrium (LD) with rs7203560: they hypothesized that this could be a marker for one or more variants in or near MCS that downregulate α-globin expression causing a mild thalassemia-like effect (Milton et al., 2013).

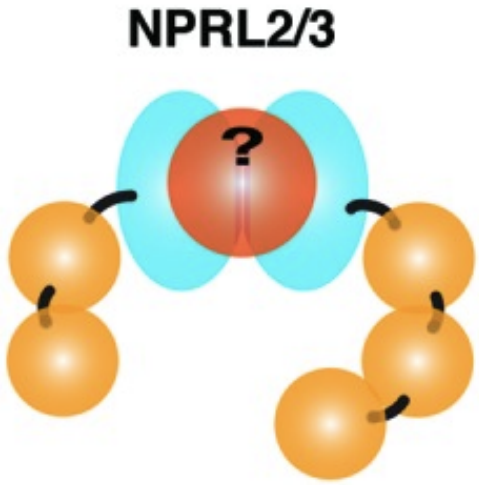

Figure 13. Levine et al. model for Nprl2/Nprl3 interaction. The longin domains (LDs) at their N-terminal (light blue) could form a platform for a small GTPase unknown (red), whereas the helix turn helix (HTH) domains at their C-terminal (orange) might bind DNA, RNA or proteins (Levine et al., 2013).

Mammalian Target of Rapamycin (mTOR)

TOR protein is the target of Rapamycin, a drug produced by Streptomyces Hygroscopicus bacteria and discovered in the 1970s (Vézina et al., 1975). The remarkable effect of this drug was its strong antiproliferative property and in the 1990s its cellular target was determined in yeast and named TOR (Cafferkey et al., 1993; Kunz et al., 1993). The mammalian homologous was then found and named mTOR (Brown et al., 1994; Sabatini et al., 1994; Sabers et al., 1995).

mTOR is a 289 kDa serine/threonine kinase that belongs to the phospho-inositide 3-kinase (PI3K)-related kinase family. It forms 2 protein complexes with different interactors, and each complex, called mTORC1 and mTORC2, has different Rapamycin sensitivity and different upstream inputs and downstream effectors.

mTORC1 and mTORC2 are two large protein complexes that in addition to the catalytic subunit have respectively five and six interacting components. These are: regulatory-associated protein of mTOR (Raptor); mammalian lethal with Sec13 protein 8 (mLST8); proline-rich AKT substrate 40 kDa (PRAS40); and DEP-domain-containing mTOR-interacting protein (Deptor).

The other components of mTORC2 are: rapamycin-insensitive companion of mTOR (Rictor); mammalian stress-activated protein kinase interacting protein (mSIN1); protein observed with Rictor-1 (Protor-1); mLST8; and Deptor (Laplante and Sabatini, 2009).

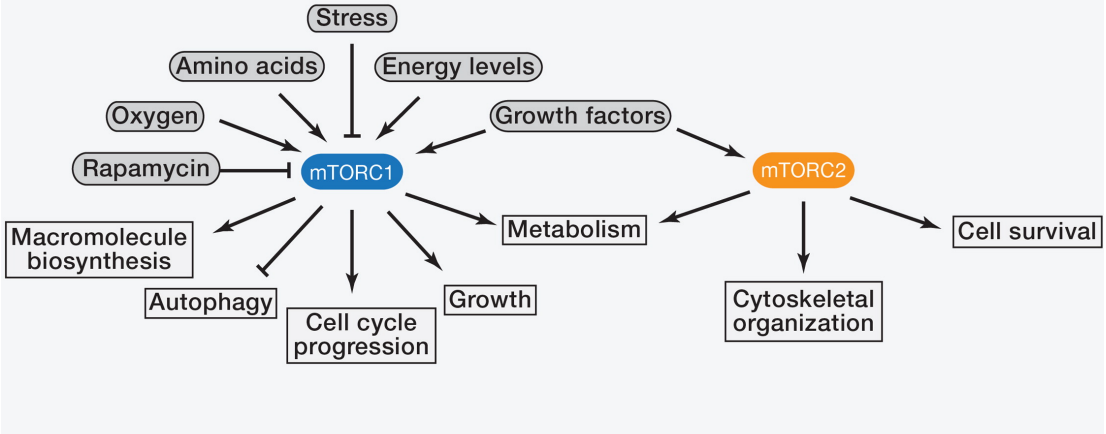

Figure 14. The two mTOR protein complexes regulate different pathways and have distinct upstream regulators. mTORC1 responds to oxygen, amino acids, stress, energy levels, growth factors and it is sensitive to acute treatments with Rapamycin. It promotes cell growth by regulating metabolism and promotes cell cycle progression. mTORC2 is sensitive to chronic Rapamycin but not to acute treatment. It regulates cell survival, metabolism and cytoskeletal organization (adapted from Laplante and Sabatini, 2012).

Up to now, mTORC1 pathway has been better characterized than mTORC2. It regulates cell growth integrating different inputs like growth factors, stress, energy status, oxygen, and amino acids and controls fundamental cellular pathways including protein and lipid synthesis and autophagy.

Amino acid sensing by mTORC1 initiates within the lysosomal lumen inducing its translocation to the lysosomal surface, where it becomes active. This activation requires a signalling machine associated with the lysosomal membrane that consists of the Rag GTPases, the Ragulator complex, and the vacuolar ATPase (v-ATPase) (Bar-Peled et al., 2013).

Nprl3 takes part of a protein complex that inhibits mTORC1

An important contribution to understand mTORC1 regulation was provided by the work of Neklesa and Davis, 2009 (in yeast) and Bar-Peled et al., 2013 (in mammals). In these papers, Nprl3 protein was proved to be part of a protein complex that inhibits mTORC1 activity.

In details, Bar-Peled et al. identified GATOR as a complex interacting with Rag GTPases present on lysosomal surface and composed of two subcomplexes called GATOR1 and 2.

Inhibition of GATOR1 subunits (DEPDC5, Nprl2, and Nprl3) makes mTORC1 signalling resistant to amino acid deprivation, whereas inhibition of GATOR2 subunits (Mios, WDR24, WDR59, Seh1L, Sec13) suppresses mTORC1 signaling. Epistasis analysis showed that GATOR2 negatively regulates DEPDC5.

GATOR1 is considered a tumor suppressor since mutations of its components are associated to hyperactive mTORC1 and therefore to a cancer phenotype: in fact, loss of DEPDC5 and Nprl2 genes was observed in human glioblastoma and ovarian cancer tissues (Bar-Peled et al., 2013).

Nprl3 ablation was studied by Kowalczyk et al., 2012 taking advantage of a mouse model in which the constitutive promoter of Nprl3 has been removed. They showed that the mice die during late gestation with cardiovascular system defects and the main expression alterations have been found in protein synthesis and cell cycle genes, because of mTOR pathway perturbation.

Phenotypically mice showed severe myocardium abnormalities suggesting that Nprl3 is a candidate gene for harbouring mutations in individuals with similar pathological symptoms (Kowalczyk et al., 2012).

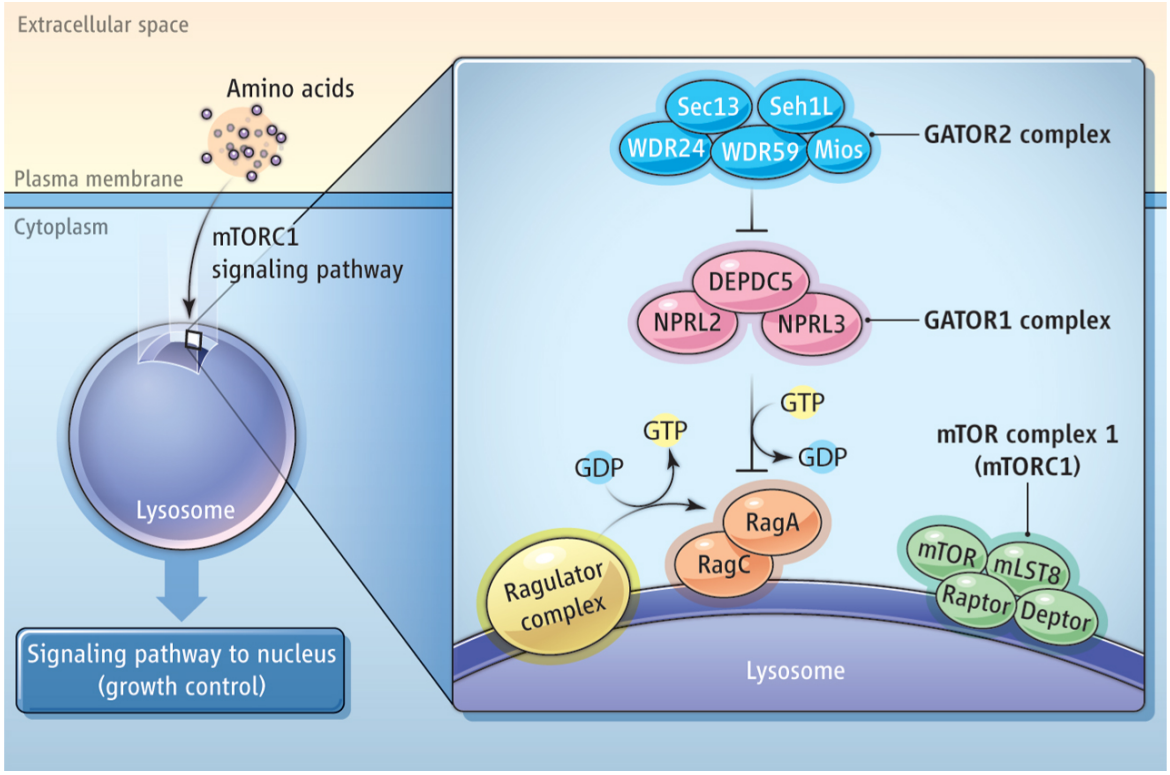

Figure 15. The GATOR complexes and their role in mTORC1 regulation. GATOR2 inhibits GATOR1 that in turn inhibits Rag GTPase A (RagA) activation. Ragulator complex has an opposite effect because it stimulates GTP loading on RagA, which activates mTORC1 in response to amino acid availability (Shaw, 2013).

mTORC1 and human diseases

In eukaryotes TOR is the major sensor of nutrients, energy and stress. Therefore, it is not surprising that alterations in its pathway have been correlated with diseases and conditions where growth and homeostasis are compromised like cancer, metabolic diseases and ageing.

mTORC1 pathway alterations have many phenotypic consequences. First, mTORC1 inhibition is correlated with increased lifespan (Bjedov et al., 2010; Hansen et al., 2008). mTORC1 pathway inactivation also counters aging and mitigate age-related diseases including neurodegenerative diseases (Johnson et al., 2013).

In this context, the mTORC1 inhibitor Rapamycin decreases the formation of toxic huntingtin aggregates in fly and mouse model of Huntington's Disease (Ravikumar et al., 2004), prevent the synthesis and aggregation of tau protein in Alzheimer's Disease *in vitro* models (Tang et al., 2013) and confers neuroprotection from PD toxins both *in vitro* and *in vivo* (Malagelada et al., 2010).

Many findings support the importance of mTOR in cancer pathogenesis. Several human cancers harbour mutations in the PI3K signalling pathway components, which is upstream both to mTORC1 and mTORC2. p53 loss, a common feature in cancer, promotes mTOR. Other tumor suppressors like Tsc1/2, serine threonine kinase 11 (Lkb1), Pten, and neurofibromatosis type 1 (Nf1) are associated with mTOR upstream regulation that is fundamental for cancer growth, survival and proliferation (Laplante and Sabatini, 2012). Figure 16 shows the downstream effects of mTOR stimulation in cancer.

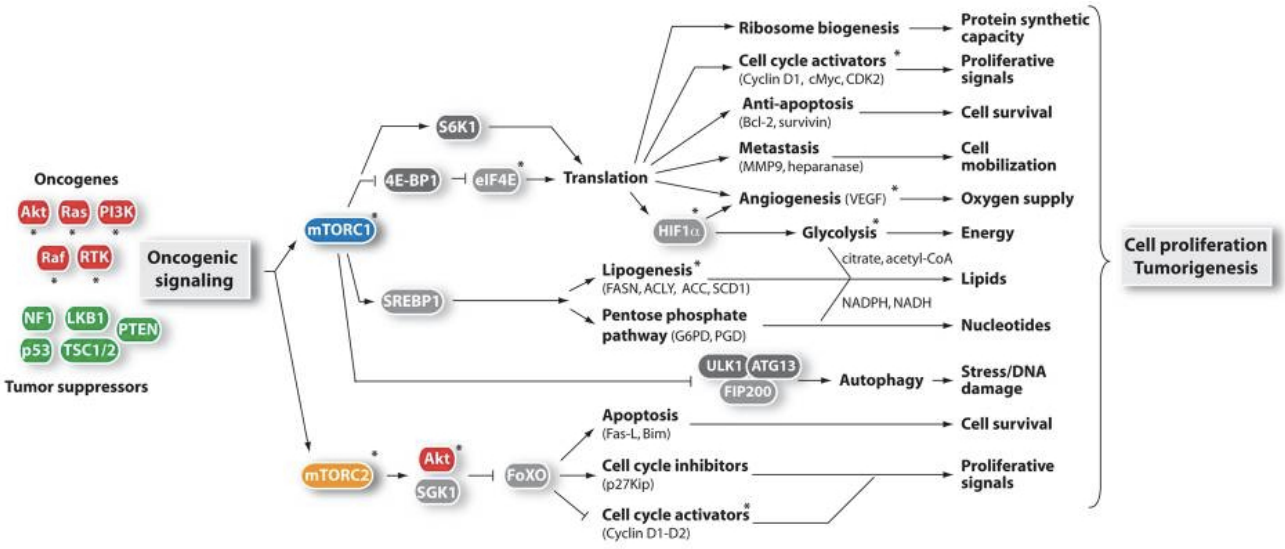

Figure 16. mTOR signalling in cancer. Tumorigenesis is promoted upon mTOR stimulation: oncogenes are in red and tumor suppressor in green. The asterisk indicates proteins currently targeted for cancer therapy (adapted from Laplante and Sabatini, 2012).

Since it responds to nutrients and growth factor levels, mTOR signalling has been also associated to metabolic diseases such as obesity, non-alcoholic fatty liver disease, insulin resistance and type 2 diabetes (Laplante and Sabatini, 2012).

In the blood, mTORC1 has been correlated with erythropoiesis, anemia and methionine homeostasis.

Knight et al., 2014 found high levels of phosphorylated ribosomal protein S6 in mouse blood reticulocytes, as a consequence of a high mTORC1 activity. They raised the hypothesis that mTORC1 can be regulated by iron availability, as shown by Ohyashiki et al., 2009, and could modulate iron metabolism, as demonstrated by Bayeva et al., 2012 and La et al., 2013. In fact, they detected a marked reduction in mTORC1 signalling in the red blood cells (RBCs) of mice in response to iron-deficient diet.

Furthermore, they demonstrated that in RBCs, mTORC1 activation induces macrocytic anemia (RBC size increase, increased hemoglobin, lower RBC number and lower hematocrit). On the contrary, genetic mTORC1 ablation induces hypochromic microcytic anemia and perinatal lethality.

Finally, in the adult mice, mTOR inhibition strongly impairs erythropoiesis influencing the growth and proliferation of erythroid progenitors (Knight et al., 2014).

Interestingly, some types of hereditary anemias can be ameliorated by pharmacological activation of mTORC1 (Jaako et al., 2012; Payne et al., 2012).

Consistent with these findings, Dutchak et al., 2015 showed that the GATOR1 component Nprl2 is required for mouse viability and that its absence significantly compromises hematopoiesis.

Furthermore, they demonstrated that Nprl2 regulates the uptake and availability of cobalamin (vitamin B12), which in turn is required for methionine and *S*-Adenosyl methionin (SAM) synthesis that are part of important metabolic pathways for hematopoiesis (Klee, 2000; Koury and Ponka, 2004).

Recently, Sim and colleagues (2015) identified Nprl3 gene mutations and subsequent mTOR activation in patients with familial focal cortical dysplasia (FCD).

Rapamycin effects on PD

Besides the importance of mTORC1 in homeostasis maintenance and the perturbation of its pathway upon disease, many works have studied the therapeutic effects of mTORC1 inhibitor Rapamycin in PD.

Two works published in 2010 show that this drug reduces dopaminergic nigrostriatal degeneration in the MPTP model of PD. They hypothesize two different mechanisms of neuroprotection: Malagelada et al. proposed that Rapamycin blocks translation of the pro-cell death protein RTP801, which is induced in affected neurons of PD patients and causes neuronal cell death by suppressing activation of mTOR (Malagelada et al., 2010). On the other hand, Dehay et al. proposed that the effect is driven by induction of lysosome-mediated autophagic degradation (Dehay et al., 2010). In particular, Rapamycin triggers lysosomal biogenesis, stimulates autophagosome-lysosome fusion and enhances lysosome-mediated clearance of accumulated autophagosomes.

Autophagic stimulation may also enhance neuronal survival by decreasing apoptosis: in PD models autophagy is impaired, leading to dysfunctional mitochondria accumulation and release of mitochondrial pro-apoptotic factors, such as cytochrome c (Bové et al., 2011). Moreover, the ubiquitine-proteasome system, altered in PD, may be affected by Rapamycin treatment (Dauer and Przedborski, 2003).

Finally, working on a mouse model, Santini and colleagues showed that levodopa stimulates mTORC1 and this sustained activity induces dyskinesia, a typical side effect that appears in PD patients medicated with levodopa. The inhibition of mTORC1 upon Rapamycin administration could ameliorate this side-effect without affecting levodopa benefits (Santini et al., 2009).

Therefore, the mechanism of neuroprotection exerted by Rapamycin seems to involve various molecular pathways and it is likely that all contribute to the neuroprotective effect on PD (Bové et al., 2011).

Preliminary data

As shown in Calligaris et al. (BMC Genomics; in press), the laboratory of Prof. Gustincich aimed to identify gene expression patterns in peripheral blood of *de novo* and drug-naïve PD patients by comparing 40 sporadic PD versus 20 HCs ("Discovery set"). To this purpose, patients were enrolled at the early clinical stage of the disease as evaluated by a neurologist. Subjects did not take any centrally acting drugs in the previous 6 months. Most patients showed a prevalent asymmetric parkinsonian symptomatology that, in the majority of cases, comprised the classical triad of tremor, bradykinesia, and rigidity. Accordingly, an asymmetric reduction of striatal activity on the ^{123}I-FP-CIT SPECT images was observed. The prevalent case consisted of a putaminal alteration contralateral to the clinical most affected side.

Blood was collected from study subjects into PAXgene Blood RNA tubes (PreAnalytiX, Hombrechtikon, CH) after a fasting period and at the same time of the day to limit circadian-dependent variability. Hybridization targets were synthesized with Ovation^{TM} Whole Blood Solution (NuGEN) and hybridized to HG−U133A 2.0 arrays (Affymetrix, Santa Clara, CA), investigating the expression of 18400 transcripts.

By applying Ranking-Principal Component Analysis, PUMA and Significance Analysis of Microarrays, gene expression profiling discriminated patients from HC and identified a common repertory of 54 genes as differentially expressed in blood. The majority of these were also present in DA neurons of the *substantia nigra*, the key site of neurodegeneration. Together with neuronal apoptosis, lymphocyte activation and mitochondrial dysfunction, already found in previous analysis of PD blood and *post-mortem* brains, they unveiled transcriptome changes enriched in biological terms related to epigenetic modifications including chromatin remodeling and methylation. Candidate transcripts were validated by RT-qPCR in an independent cohort of 12 patients and controls ("Validation set").

Aim of this work

In this work we took advantage of nanoCAGE technology to find alterations in the blood transcriptome of 20 drug naïve de novo PD patients and 20 HC.

The most up-regulated nanoCAGE tag is located in the third intron of the gene Nitrogen Permease Regulator Like Protein 3 (Nprl3).

The aim of my PhD thesis was to identify the full-length transcript associated to the nanoCAGE tag, validate it, and to test whether it may represent a peripheral biomarker of PD.

Here I show that the tag represents an alternative Transcription Start Site of Nprl3 (TagNprl3). It is associated to a TCT motif (YC+1TYTYY) for initiation of transcription, which has been found to be specific for ribosomal protein-coding genes and those involved in protein synthesis. The tag maps to a 29nt minisatellite that is found repeated 16 times in the reference genome. High tag expression is associated to an allelic variant of 13 repeats. To our knowledge, this is the first time that a minisatellite variant is both a TSS and an eQTL.

Unfortunately, high TagNprl3 expression resulted not to be correlated to PD but to heterozygosity. Furthermore, allelic frequencies were not correlated to PD.

We then showed that TagNprl3 is expressed in RBCs both at mRNA and protein levels giving rise to an isoform truncated at the N-terminal. This is able to interact with its protein partner Nprl2 and its overexpression inhibits cell proliferation.

This work provides hints for Nprl3 protein function in blood and may suggest a testable hypothesis linking mTOR activity to genomic polymorphisms in modifier genes.

Materials and methods

Blood collection and RNA purification

The study was approved by the local institutional Ethical Committee at the Movement Disorders Center of the Neurologic Clinic of Trieste (Italy). Study participants gave written informed consent. We enrolled 20 patients with a first clinical diagnosis of PD, according to the UK Parkinson's Disease Society Brain Bank criteria. Twenty healthy age- and ethnicity-matched control subjects (HC) travelling with the patients were also included in the study.

Blood was collected from study subjects after a fasting period. Samples were harvested directly and sequentially into 8 PAXgene Blood RNA tubes (PreAnalytiX, Hombrechtikon, CH) via a 21−gauge butterfly needle and then frozen and kept at -80°C. Total RNA was purified using PAXgene^{TM} Blood RNA kit (PreAnalytiX GmbH, Qiagen, Hilden, Germany) and DNaseI treatment was performed by 'on-column' treatment as recommended by manufacturer's instructions plus a second treatment subsequent to elution. RNA was then purified using RNase column (Qiagen, Hilden, Germany) and quantified by Nanodrop ND-100 Spectophotometer (NanoDrop Technologies; Wilmington, DE). RNA integrity was determined with 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) and exclusively samples with RIN $\geq$ 8 were included in the subsequent investigations.

These steps were performed by Dr. Calligaris in the laboratory of Prof. Gustincich.

Nano CAGE

For a detailed description of nanoCAGE please refer to Plessy et al., 2010. This step was performed by Dr. Dave Tang in the laboratory of Prof. Carninci (Riken, Japan).

Rapid amplification of cDNA ends (RACE)

RACE technique allows the amplification of full-length 5′ and 3′ ends of cDNA starting from a known partial sequence obtained for example from library screening

such as nanoCAGE.

RACE was performed using the GeneRacer cDNA amplification kit (Invitrogen).
1 µg of total RNA has been retrotranscribed into cDNA modified with the GeneRacer 5' and 3' adaptors following the manufacturer's protocol, using poly-T primer in the 20 µl reverse transcription reaction.
All RACE reactions were performed by nested PCR. For the first amplification, 1 µl cDNA obtained from the retrotranscription reaction was used as a template together with GeneRacer™ 5′ Primer or the GeneRacer™ 3′ Primer, when performing 5' or 3' RACE reactions, respectively, and gene-specific primers: REV_first_5RACE for 5'RACE and FW_3RACE for 3'RACE (see Table 2).
PCR was performed with Platinum Taq High Fidelity (Invitrogen) under the following conditions: 94°C for 2 min, followed by 5 cycles at 94°C for 30 s, 72°C for 1 min; 5 cycles at 94°C for 30s, 70°C for 1 min, 25 cycles at 94°C for 30s, 68°C for 30 s, 68°C for 1 min, and a final extension at 68°C for 10 min.
1 µl of this PCR product was used as template for the nested PCR together with GeneRacer™ 5′ Nested Primer or the GeneRacer™ 3′ Nested Primer, when performing 5' or 3' reactions, respectively, and a second nested gene-specific primers: REV_nested_5RACE for 5'RACE and FW_3RACE (the same of the first reaction) for 3'RACE.
Nested PCR was performed under the following conditions: 94°C for 2 min, followed by 40 cycles at 94°C for 30s, 65°C for 30 s, 68°C for 2 min, and a final extension at 68°C for 10 min. PCR products were resolved on 1% agarose gel. The amplicons were cloned in pGEM-T Easy vector (Promega) and sequenced at the Eurofins MWG Operon Inc. facility.

Primer	**Sequence (5' → 3')**
GeneRacer™ 5′ Primer	CGACTGGAGCACGAGGACACTGA
GeneRacer™ 3′ Primer	GCTGTCAACGATACGCTACGTAACG
GeneRacer™ 5′ Nested Primer	GGACACTGACATGGACTGAAGGAGTA
GeneRacer™ 3′ Nested Primer	CGCTACGTAACGGCATGACAGTG
REV_first_5RACE	CCCAGAGCATGCTGTAGCAGTGTT
FW_3RACE	GAGTGTGTGATCCTGTTTCTCAGCGTG
REV_nested_5RACE	ATAACATCTGAAAACCAGACAAGAAACA

Table 2. List of primers used in RACE assays.

Identification of Minisatellite Polymorphism

Genomic DNA has been extracted from blood samples taking advantage of QIAamp DNA Blood Mini Kit (Qiagen) following the manufacturer's protocol (Dr. Calligaris, Dr. Vlachouli, Dr. Finaurini).

PCRs were performed in 96 well plates, on a total of 20 ng/5 µl of human genomic DNA per well, by using ExTaq Takara on a total of 50 µl reaction volume, as specified by the manufacturer. Primers were used at a final concentration of 200 nM and are indicated on Table 3.

The two primers were designed to map externally to the repeated region in order to obtain the amplification of the entire minisatellite. The cycling conditions were: 94°C for 3 min, followed by 35 cycles at 94°C for 30 s, 62.4°C for 30 s, 72°C for 1 min, and a final extension at 72°C for 5 min. Reaction products were visualized on a 1.3% agarose gel.

Primer	**Sequence (5' → 3')**
VNTR_FW	GCAGAAGTGCCACCATTAAGCA
VNTR_REV	AAACCCCATGGTAAGCGTTGA

Table 3. Primers used for minisatellite polymorphism identification.

Luciferase reporter assay

This assay allows testing genomic sequences for their putative function in the regulation of gene expression.

It gives a quantitative measurement of the level of expression of the luciferase reporter gene under the control of the genomic DNA sequence to be tested.

Minisatellite alleles have been cloned into pTAL-Luc vector (Clontech) that is an enhancerless vector that express firefly Luciferase gene under the control of a minimal TATA-like promoter region from the Herpes simplex virus thymidine kinase (HSV-TK) promoter and it is ideal for the study of putative enhancer sequences.

The inserts have been amplified with modified Table 3 primers, adding XhoI and KpnI restriction sites in order to obtain four different constructs, with the two minisatellite alleles in both orientation, since enhancer activity might be orientation-dependent.

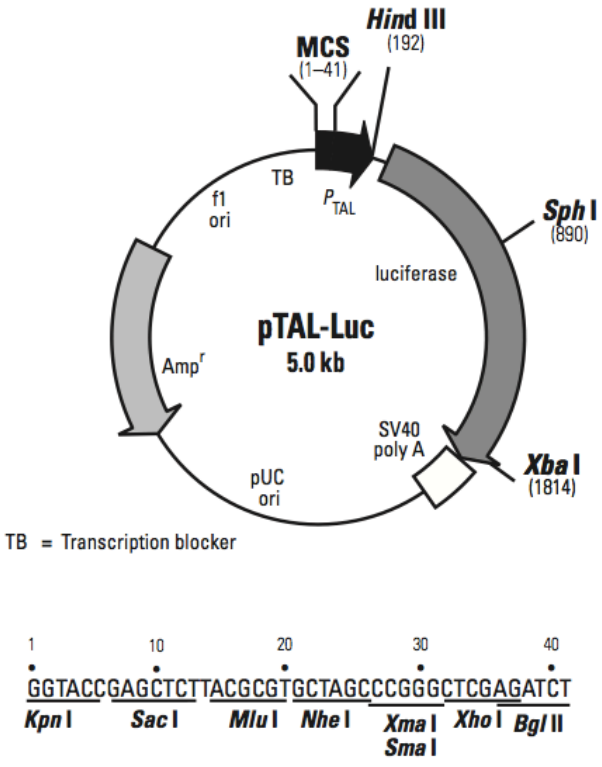

Figure 17. Restriction map and multiple cloning site (MCS) of pTal-Luc.

The assay was performed taking advantage of lymphoblastoid K562 cell line (chronic myelogenous leukemia) maintained in RPMI 1640 (GIBCO) supplemented

with 10% fetal bovine serum (SIGMA-ALDRICH), 1% GlutaMAX, 100 U/ml penicillin and 100μm/ml streptomycin (SIGMA) at 37°C in a humidified CO_2 incubator.

K562 cells were transfected with Lipofectamine 2000 Transfection Reagent (Invitrogen) following manufacturer instructions. For each well the pTAL-Luc vector to be tested was transfected together with Tk-Ren normalization vector (Promega), that express renilla luciferase under the same TATA-like promoter, in a ratio 10:1.

After 48 hours from transfection, the activities of firefly and renilla luciferase were measured using the Dual-Luciferase Assay Kit (Promega).

Luciferase activity ratio (firefly/renilla) for each transfection condition was normalized to the ratio obtained with control plasmid (pTAL-Luc empty). Normalization to the mean activity of the control condition gave the fold-change in luciferase activity values presented.

Genome wide association study (GWAS)

We carried out a genome-wide association study (GWAS) to identify hypothetical variants associated to minisatellite polymorphism.

The analysis was carried out by Prof. Pio D'Adamo (Consorzio per il Centro di Biomedicina Molecolare – CBM, Trieste) on the genomic DNA of PD and HC taking advantage of HumanOmniExpressExome BeadChip (Illumina) array.

The array is designed to capture the greatest amount of common SNP variation.

RT-PCR and Real-Time TaqMan PCR

To validate the full-length TagNprl3 transcript, 1 μg of total RNA was reverse-transcribed in a final volume of 20 μL using Superscript III Reverse Transcriptase (Invitrogen), 25 ng random hexamers and 2.5 μM oligo(dT) 20 primers according to the manufacturer's recommendations.

PCR mix was prepared by adding 1 μL of the so prepared cDNA to 5 units of TaKaRa Ex Taq, 10X Ex Taq Buffer, dNTP Mixture (200 μM final concentration each) (TaKaRa), 200 nM of primer FW_3RACE (see Table 2), 200 nM of primer REV (5'-CCTTTCCAAACCTGCGCACC-3'), and water to a final volume of 50 μL.

The FW_3RACE is the same forward primer used to amplify the full-length transcript in 3'RACE assay, while the REV primer was designed in the 3'UTR region of TagNprl3. The couple of primers used to amplify Nprl3 isoform 1 is indicated on Table 4.

The PCR protocol is the following: 95°C for 5 min, followed by 40 cycles at 95°C for 15 s, 60°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 5 min. PCR products were resolved on 1% agarose gel.

Real-Time PCR is a technique that allows determining the starting quantity of a template through amplification and detection of nucleic acids. It is an accurate, specific and high sensitive method that, by the use of fluorescent molecules, such as DNA-binding dyes or fluorescently labeled sequence-specific probes, detects the amplicon at each step of amplification. The introduction of TaqMan fluorescently labeled sequence-specific probes increases the specificity of the signal and the reproducibility of the assay.

Real-time PCR was performed in the presence of 1 μL of cDNA template, TaqMan gene expression master mix, commercially available TaqMan gene expression assay for PGK1 (Hs99999906_m1), 250 nM specific primer mix, 150 nM specific TaqMan MGB probe (6-carboxyfluorescein (FAM) dye-labeled) and run on an iCycler IQ (Bio-Rad) in a 20 μL reaction volume according to the manufacturer's instructions. PGK1 has been used as a reference gene for this work because it was among the most reliable reference genes for peripheral blood gene expression analyses (Calligaris et al., BMC Genomics; in press).

The specific primers and probes have been designed on Nprl3 isoform 1 and on TagNprl3 and are indicated on Table 4 and on Figure 18.

Thermal cycler conditions were as follows: 95°C for 10 min, followed by 40 cycles of amplification at 95°C for 15 s and 60°C for 1 min.

Standard curves of cDNA ranging from 120 ng to 0.2 ng were used to verify that the 50 ng dilution tested was within the linear range of reaction. Primer efficiency and multiplexing efficacy was verified by linear regression to the standard curve with a slope near -3.32 representing acceptable amplification efficiency.

The amplified products were separated on a 2% agarose gel. Results were normalized to PGK1 and the initial amount of the template of each sample was determined as relative expression versus a pool of healthy control samples used as

calibrator. The relative expression of each sample was calculated by the formula 2exp-ΔΔCt (User Bulletin 2 of the ABI Prism 7700 Sequence Detection System).

Primer (5' → 3')	**Nprl3 assay**	**TagNprl3 assay**
FW	CCCCACGGCGGGATGCGGGA	GTGATCCTGTTTCTTGTCTGGT
REV	TCGCCCGTGTTGCTGGCAGCGT	CAGAGCATGCTGTAGCAGTGT
probe	AGCCAGGAGCACCCGGCGTCC	GTTATTCTGGCAACAATTTTGGCA

Table 4. List of primers designed for targeting Nprl3 and TagNprl3 and used in the Real-Time TaqMan PCR assays.

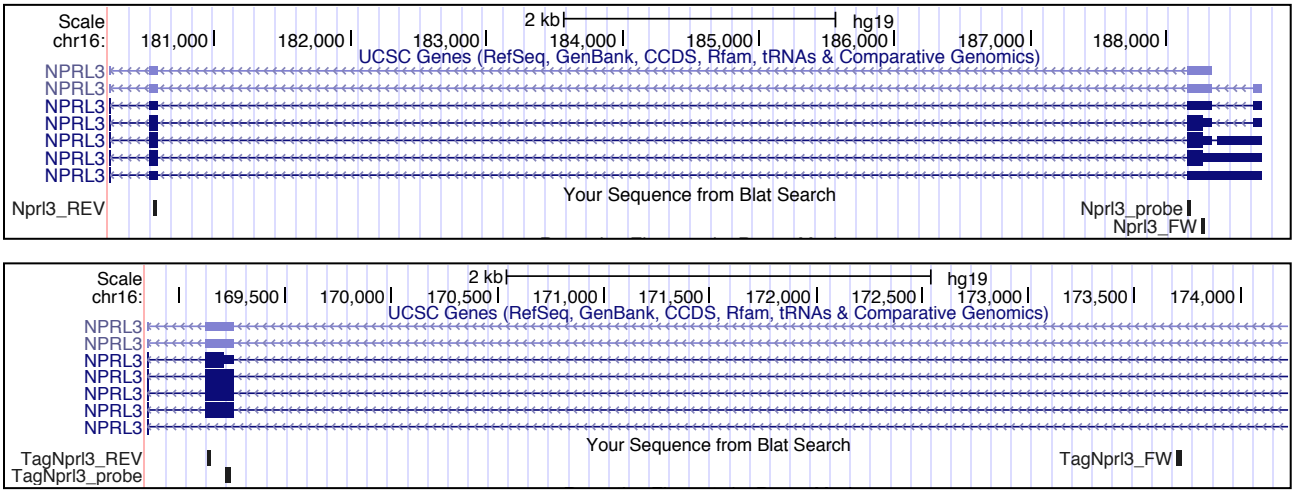

Figure 18. UCSC Genome Browser visualization of Nprl3 and TagNprl3 TaqMan assays.
Upper panel: Nprl3 assay designed on second and third exon of Nprl3 isoform 1.
Lower panel: TagNprl3 assay designed on intronic minisatellite and fourth Nprl3 exon.

RNA and protein preparation from blood fractions

Blood was collected into EDTA tubes and the plasma, the red blood cells and the peripheral blood mononuclear cells (PBMC) have been separated starting from the whole blood samples by using Ficoll Histopaque gradient following the manufacturer's protocol (SIGMA-ALDRICH).

To extract total RNA from the blood fractions each sample was added with a proper amount of TRIzol reagent (Invitrogen) followed by manufacturer's protocol.

A fraction of the total RNA sample was treated with DNase I (Ambion) at 37°C for 1 hour, and the sample was then purified on RNAeasy mini kit columns (Qiagen). The

final quality of RNA sample was tested on the Agilent 2100 bioanalyzer using the Eukaryote Total RNA Nano assay.

The PBMC protein lysate has been prepared adding a proper amount of 2X SDS sample buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) sodium dodecyl sulfate, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, 200 mM β-mercaptoethanol). The protein lysates from whole blood, plasma and red blood cells have been prepared according to Lin et al., 2012 ("Whole blood sample preparation" chapter) followed by an albumin depletion using trichloroacetic acid (TCA)/acetone (Chen et al., 2005).

The obtained precipitates were diluted in a proper amount of 2X SDS sample buffer.

Expression plasmids

Nprl3: pcDNA3 vector containing the full-length DNA sequence of human Nprl3 isoform 1 and pcDNA3 vector containing coding DNA sequence of human Nprl3 isoform 2 were kindly provided by Licio Collavin's laboratory (Consorzio Interuniversitario per le Biotecnologie (C.I.B.), Trieste).

TagNprl3: the full-length DNA sequence of TagNprl3 was amplified from whole blood cDNA modified with the GeneRacer 5' and 3' adaptors via PCR with the GeneRacer^{TM} 5′ Primer (see Table 2) and a gene specific reverse primer: 5'-TTTTGCTTGGCCTGGCTTTTATCTTGAA-3'.

The pGEM T-easy-TagNprl3 plasmid was generated by subcloning the cDNA sequence into the EcoRI cloning site of the pGEM T-easy cloning vector from Promega for sequencing. This cassette was finally cloned into the pcDNA3.1 (-) expression vector using the same restriction enzymes.

Nprl2_Myc: pcDNA3.1- vector containing the coding region of human Nprl2 amplified from whole blood cDNA. The Myc tag was fused to the C-terminus of the protein.

Cell culture, transfections and immunoblotting

HEK 293T (human embryonic kidney) cells were grown in DMEM (GIBCO) supplemented with 10% fetal bovine serum (SIGMA-ALDRICH), 100 U/ml penicillin and 100μm/ml streptomycin (SIGMA) at 37°C in a humidified CO_2

incubator. HEK cells were transfected with FuGENE HD Transfection Reagent (Promega) following manufacturer instructions. After 48 hours from transfection, cells were either collected with 2X SDS sample buffer or analyzed via immunofluorescence.

All the protein lysates have been boiled 5 min at 95°C and analyzed by Western Blot. For Western Blot, samples were resolved on SDS/PAGE, and proteins were transferred to nitrocellulose membrane (Amersham, GE Healthcare). Membrane was blocked with 5% non-fat milk in Tris buffer saline solution (TBST), then incubated overnight at 4°C with the following primary antibodies: anti-Nprl3 (ab121346, Abcam), anti-β actin (A5441, SIGMA-ALDRICH).

Proteins were detected by horseradish peroxidase-conjugated secondary antibodies (DakoCytomation, Glostrup, Denmark) and ECL reagents (Amersham, GE Healthcare).

Immunofluorescence

For immunocytochemistry experiments, cells were fixed in 4% paraformaldehyde (SIGMA ALDRICH) for 10 minutes, then washed with phosphate-buffered salin solution (PBS) two times, treated with 0.1M glycine for 4 minutes in PBS and permeabilized with 0.1% Triton X-100 in PBS for another 4 minutes. After washing with PBS and blocking with 0.2% BSA, 1% NGS, 0.1% Triton X-100 in PBS (blocking solution), cells were incubated with the indicated antibodies diluted in blocking solution for 90 minutes at room temperature. After washes in PBS, cells were incubated with labelled secondary antibodies for 60 minutes. For nuclear staining, cells were incubated with 1μg/ml DAPI for 5 minutes. Cells were washed and mounted with Vectashield mounting medium (Vector). All images were collected using a Leica DM6000 fluorescence microscope.

Co-immunoprecipitation

HEK cells were transfected with FuGENE HD Transfection Reagent (Promega) following manufacturer instructions. After 24 hours from transfection, cells were harvested and lysed in the following buffer supplemented with cOmplete EDTA-free

Protease Inhibitor Tablets (Roche): 150 mM NaCl, 50 mM Tris HCl pH 7.4, 0.5% NP-40, 1mM EDTA.

After 30 minutes of lysis at 4°C samples were sonicated twice (7 micron amplitude, 10 seconds) and then centrifuged 30 minutes, 4°C at maximum speed. Supernatants were collected and a fraction was saved as input. Lysated were incubated with 1 µl of Myc-Tag (9B11) antibody from Cell Signaling. After 3 hours of incubation, 40 µl of Protein G Sepharose beads (GE Healthcare) were added to each sample and incubated 1 hour. After washing, beads were dried and eluted in 2X SDS sample buffer, boiled 5 min at 95°C and analyzed by Western Blot.

Bromodeoxyuridine (BrdU) assay

Bromodeoxyuridine is a synthetic nucleoside analog of thymidine. BrdU is incorporated in living cells during DNA replication (S phase of the cell cycle) and it is commonly used to detect proliferating cells.

HEK cells were transfected with FuGENE HD Transfection Reagent (Promega) following manufacturer instructions. As a control, a pcDNA3.1- vector expressing GFP was used. After 48 hours from transfection BrdU was added to each plate at a final concentration of 15 µg/ml for 1 hour.

Cells were then fixed in 4% paraformaldehyde (SIGMA ALDRICH) for 10 minutes, then washed with phosphate-buffered salin solution (PBS) two times, treated with 0.1M glycine for 4 minutes in PBS and permeabilized with cold acetone for 30 seconds. After washing with PBS cells are treated with HCl 3N for 15 minutes to separate DNA into single strands so the primary antibody can access the incorporated BrdU. After several washes with PBS cells were incubated at 37°C with a blocking solution of 1% BSA for 30 minutes, followed by 37°C incubation with an anti-BrdU antibody (1:50, SIGMA ALDRICH) and an anti-Nprl3 antibody (1:500, Abcam) in blocking solution for 90 minutes. After washes in PBS, cells were incubated with labelled secondary antibodies for 60 minutes. For nuclear staining, cells were incubated with 1µg/ml DAPI for 5 minutes. Cells were washed and mounted with Vectashield mounting medium (Vector). All images were collected using a Leica DM6000 fluorescence microscope.

Results

Blood transcriptomics of PD patients with nanoCAGE

Thanks to the collaboration with the Director of Clinica Neurologica in Trieste, Prof. Gilberto Pizzolato, we collected the peripheral blood of 20 drug-naïve de novo PD patients and the peripheral blood of 20 healthy controls (HC) age and sex match.

In order to analyze the blood transcriptome of the PD samples and compare it to the HC we performed nanoCAGE in collaboration with Prof. Piero Carninci's laboratory in RIKEN Institute (Japan).

Total RNA was purified using PAXgene^{TM} Blood RNA kit (PreAnalytiX GmbH, Qiagen, Hilden, Germany) and underwent two DNaseI digestion steps. Only high quality RNAs (RIN $\geq$ 8) were included in the study as determined with 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA).

The bioinformatic analysis carried out by Dave Tang, PhD student in Carninci's laboratory, revealed a nanoCAGE tag (27 nt) that was found upregulated by more than 50 fold in 30% of PD respect to the remaining patients and HC (0%) controls.

This tag maps to the third intron of the Nitrogen Permease Regulator-like 3 (Nprl3) gene, on the telomeric region of chromosome 16, with coordinates 173660-174239 (genomic assembly hg19). The multi-mapping nature of this tag is appreciable on Figure 19, lower panel suggesting it represents a repetitive region. By Genome Browser sequence analysis, we have found that this tag maps on a minisatellite sequence of 29 nucleotides repeated 16 times. The mapping is on the (-) strand of genomic DNA, in the same direction of Nprl3 transcription. This minisatellite is present only in this genome locus and it is human-specific. The most common tag sequence is GAGTGTGTGATCCTGTTTCTCAGCGTG.

Table 5 reports the tag per million (tpm) values of the 40 nanoCAGE samples. Samples presenting the highest tag levels are PD2 (30.08), PD4 (23.12), PD7 (19.42), PD11 (46.05), PD13 (50.88), PD18 (45.57).

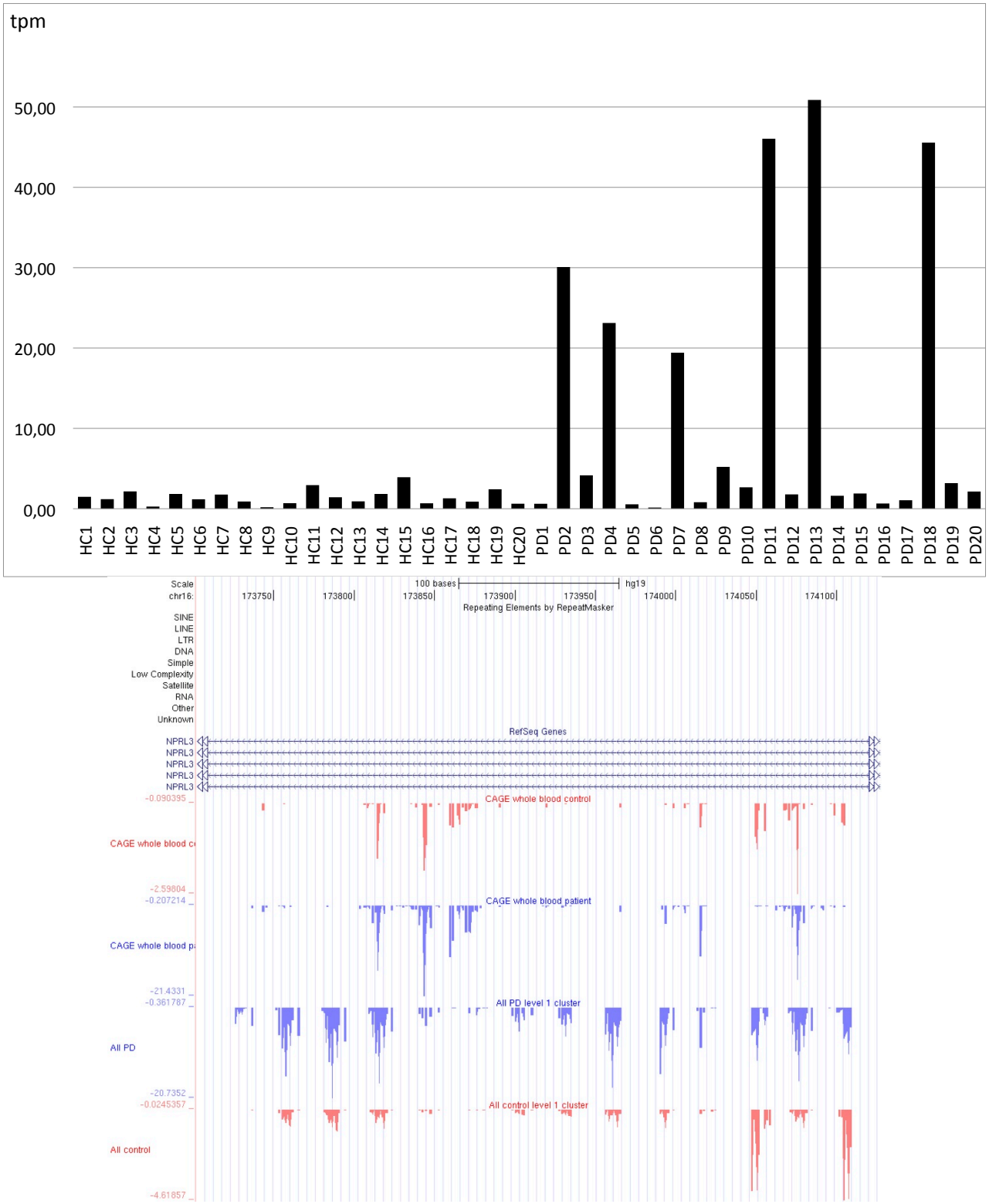

Figure 19. NanoCAGE-upregulated tag in PD.
Upper panel: the histogram shows the tag per million (tpm) value (y-axis) of the 40 nanoCAGE samples (x-axis) for chr16:173660-174239 region (hg19). High tag-expressing samples are PD2, PD4, PD7, PD11, PD13, PD18.
Lower panel: Genome Browser representation of chr16:173660-174239 region (hg19). RefSeq genes are indicated above, showing Nprl3 intronic portion, while patients (blue) and control (red) tags are mapped below. The multi-mapping tag suggests this is a repetitive region.

Sample	tpm
HC1	1.48
HC2	1.19
HC3	2.15
HC4	0.27
HC5	1.84
HC6	1.18
HC7	1.76
HC8	0.90
HC9	0.19
HC10	0.69
HC11	2.94
HC12	1.43
HC13	0.92
HC14	1.84
HC15	3.92
HC16	0.68
HC17	1.30
HC18	0.88
HC19	2.42
HC20	0.62
PD1	0.62
PD2	30.08
PD3	4.16
PD4	23.12
PD5	0.55
PD6	0.15
PD7	19.42
PD8	0.81
PD9	5.21
PD10	2.67
PD11	46.05
PD12	1.78
PD13	50.88
PD14	1.61
PD15	1.89
PD16	0.65
PD17	1.06
PD18	45.57
PD19	3.19
PD20	2.14

Table 5. NanoCAGE upregulated tag in PD: tag per million (tpm) values for each HC and PD sample, mapping on chr16:173660-174239 region (hg19).

GCGTGGGGAGCGTGTGACCCTGTTTCTCA
GCGTGGGGAGTGTGTGACCCTGTTTCTCA
CATGTTGGGAGTGTGTGATCTGTTTCTCA
GCGTGGGGAGTGTGTGATCCTGTTTCTCA
GCGTGGGGAGTGTGTGATCTGTTTCTCA
GCGTGGGGAGGGTGTGACCCTGTTTCTCA
GCGTGGGGAGTGTGTGATCCTGTTTCTCA
GCGGTGGGGAGTGTGTGATCCTGTTTCTCAAC
GTGGGGAGTGTGTGACCCTGTTTCTCA
GCGTGGGGAGTGTGTGACCCTGTTTCTCA
GCGTGGGGAGTGTGTGACCCTCTTTCTCA
GCGTGGGGAGTGTGTGATCCTGTTTCTCA
GCGTGGGGAGTGTGTGATCCTGTTTCTCA
GCGTGGGGAGTGTGTGACCCTGTTTCTCA
GCGTGGGGAGTGTGTGACCCTGTTTCTCA
GCGTGGGGAGTGTGTGATCCTGTTTCTT

Figure 20. NanoCAGE tag mapping on minisatellite genomic sequence. This is the minisatellite sequence on the minus strand of chr16:173708-174170 region (hg19). The tag GAGTGTGTGATCCTGTTTCTCAGCGTG is indicated in red and it maps 3 times.

Variable Number Tandem Repeats at the TSS

I decided to study in details the minisatellite genomic sequence to unveil the molecular basis of the differential expression between high- and low-tag individuals.

I designed two primers that map externally to the repeated region to obtain the amplification of the entire minisatellite from genomic DNAs. Dr. Vlachouli then carried out the PCR analysis of PD and HC samples of the same cohort of subjects used for nanoCAGE, revealing the presence of two agarose gel bands in a subset of individuals, corresponding presumably to a genetic polymorphism (Figure 21 upper panel).

By sequencing these amplicons, we found that the larger fragment consists of 16 repeats of the 29mer while the shorter fragment consists of 13 repeats of the 29mer. More in detail, among the 20 PD patients, 6 subjects (30%) are 16/13 heterozygous. Interestingly, these are the ones expressing high-tag counts. Among the 20 HC 1 subject (5%) is 16/13 heterozygous (sample HC15).

Figure 21 shows the minisatellite genotype of a high-tag 16/13 (PD13) and of a low-tag 16/16 (HC10). The schematic representation indicates the three repeats missing in allele 13 compared to allele 16.

We asked ourselves if there is a correlation between high-tag expression, heterozygosity and PD incidence, so we decided to analyse the minisatellite genotype on new samples from sporadic PD patients and HC, kindly provided by Clinica Neurologica (Triest) and by Associazione Donatori di Sangue (Triest).

Interestingly, heterozygosity was more prevalent in patients (8 out of 38, about 21%) compared to the controls (4 out of 60, about 7%) and to confirm this data we performed a larger screening taking 526 sporadic PD patients and 598 HC, kindly provided by the Stefano Goldwurm laboratory from Centro Studi Parkinson (Milan) and from Associazione Donatori di Sangue (Triest).

In particular, among PD patients, 101 subjects are 16/13 (19.2%); 7 subjects are 13/13 (1.3%) and 418 subjects are 16/16 (79.5%). The frequency of the allele 13 in PD was 10.9%. Among HC, 111 subjects are 16/13, (18.5%); 4 subjects are 13/13 (0.7%) and 483 subjects are 16/16 (80.8%). The frequency of the allele 13 in HC was 9.9%.

In summary no differences were observed in heterozygosity and allelic frequency between PD patients and HC.

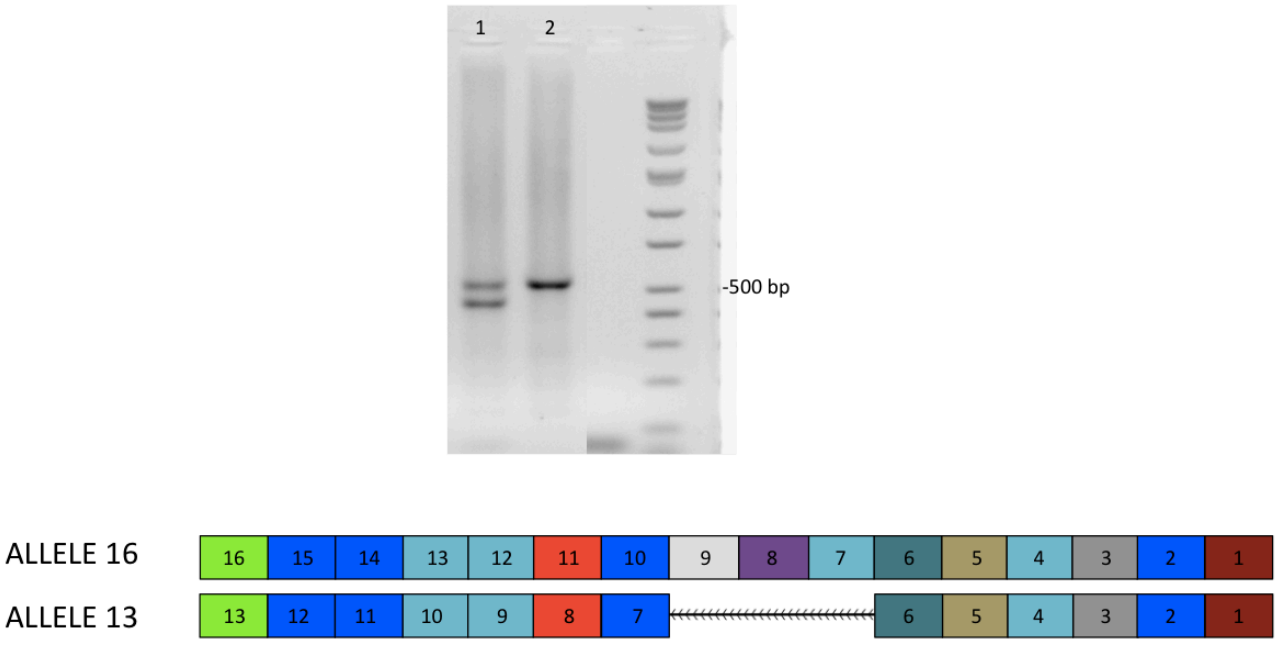

Figure 21. Genomic PCR amplification reveals that this minisatellite has two allelic forms. The upper band corresponds to 16 repeats of the 29mer, while the loIr band to 13 repeats. PD13 individual (lane 1) is 16/13 heterozygous while HC10 individual (lane 2) is 16/16 homozygous. A schematic representation of the two minisatellite alleles is presented below: repeated blocks are numbered 1-16 or 1-13 from the right to the left, since the tag mapped on the (-) strand of DNA. Same colour means same block sequence. Allele 13 lacks three repeats, corresponding to repeats 7-8-9 of allele 16.

Luciferase enhancer assay

An enhancer is a genomic sequence that increases gene transcription by recruiting diverse proteins without necessarily being in close proximity to the influenced gene.

Based on their relative genomic position to their associated gene, enhancers can be proximal or distal.

Recently, it has been demonstrated that RNA polymerase II can bind a distal enhancer producing enhancer-associated transcripts called eRNA (Kim et al., 2010; Wang et al., 2011).

Since minisatellite genotyping allowed us to assess that heterozygosity is linked to high-tag expression in our nanoCAGE samples, I hypothesized that the minisatellite allele 13 could be an enhancer and the tag-associated transcript an eRNA.

To address this model, I carried out an enhancer assay taking advantage of pTAL-Luc vector that expresses firefly luciferase gene under the control of a minimal TATA-like promoter. The expression increases when an enhancer sequence is inserted in the multiple cloning site upstream to the promoter.

I tested minisatellite allele 16 and allele 13 in both orientations because enhancer activity could be orientation-independent. As a basal control I used an empty pTAL-Luc vector that expresses minimal levels of luciferase while to normalize each transfected well I cotransfected a renilla luciferase-expressing vector in a 10:1 ratio. The peak emission of renilla luciferase is 480 nm while for firefly luciferase is 560 nm. The assay was repeated 4 times.

Relative firefly luciferase values were highly variable amongst the replicas and there is not any significant difference neither between allele 16 and allele 13 nor between empty vector and none of the minisatellite constructs. I concluded that minisatellite sequence alone could not act as an enhancer.

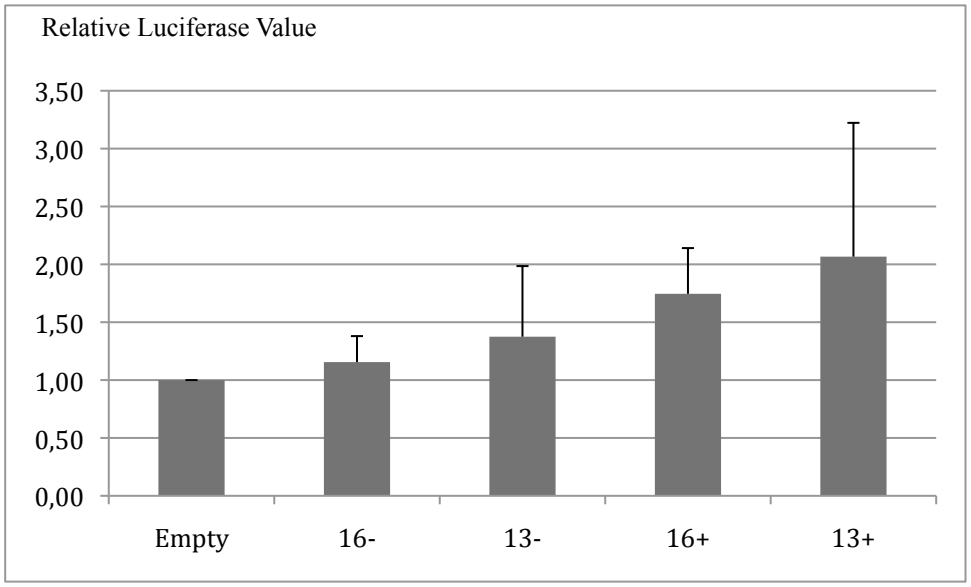

Figure 22. Relative luciferase values of enhancer assay.

Minisatellite sequences of both alleles have been cloned into pTAL-Luc vector in both orientation (- and + indicate the genomic strand) and assayed for enhancer activity. None of them could significatively enhance firefly luciferase activity and there isn't a difference between allele 16 and 13. Firefly luciferase values were normalized on renilla luciferase values.

Genome wide association study (GWAS)

The minisatellite allelic variant carrying 13 repeats has been associated in the nanoCAGE assay with high-tag expression, therefore this allele can be defined as an eQTL, a locus that modulates a quantifiable phenotypic trait.

Since the presence of allele 13 is not sufficient to enhance transcription as seen in the luciferase assay, I hypothesize that this variant could be associated in linkage disequilibrium (LD) with other loci that would contribute to the phenotypic output. LD is generally caused by interactions between genes, genetic linkage and the rate of recombination, random drift or non-random mating and population structure.

Genome-wide association studies (GWASs) explore common genetic variation in an unbiased manner: the array can capture million of SNPs distributed across the genome to find common variations in a selected pool of samples. The power of detection depends on the sample size and on the strength of the association.

I carried out a GWAS to look for variants associated to the minisatellite polymorphism. Moreover, this study allows us to discover eventual putative copy-number variations (CNVs) in the genomes of individuals.

CNVs are variations in the number of copies of selected portions of genomic DNA and are found to be very common in human population. The CNV studies become popular since they can account for gene expression levels and, in general, human phenotypic variation that might be related to a disease risk (McCarroll, 2008).

The experiment and the subsequent bioinformatic analysis was carried out as a service by Dr. Pio D'Adamo (Consorzio per il Centro di Biomedicina Molecolare – CBM, Trieste) on the genomic DNA of PD and HC taking advantage of HumanOmniExpressExome BeadChip (Illumina) array. The array is designed to capture the greatest amount of common SNP variation.

I selected 13 individuals from the nanoCAGE assay, subdividing them accordingly to the disease/healthy condition and the minisatellite genotype (Table 6).

Given that only an HC is heterozygous in this pool of samples, I needed to add three 16/13 healthy individuals that I selected from the cohort obtained from the

Associazione Donatori di Sangue (Triest), to compare four 16/16 against four 16/13 in both PD patients and controls.

Name	Condition	Genotype	Gender
4PD	PD	16/13	F
11PD	PD	16/13	F
13PD	PD	16/13	M
18PD	PD	16/13	M
5PD	PD	16/16	M
19PD	PD	16/16	F
10PD	PD	16/16	F
14PD	PD	16/16	M
15HC	HC	16/13	F
12CD	HC	16/13	F
28CD	HC	16/13	M
43CD	HC	16/13	M
2HC	HC	16/16	F
5HC	HC	16/16	M
10HC	HC	16/16	M
14HC	HC	16/16	F

Table 6. GWAS samples.
I selected 13 individuals from the nanoCAGE analysis: 8 PD (4 16/13 and 4 16/16) and 5 HC (1 16/13 and 4 16/16). The missing 16/13 HC were selected from the minisatellite genotype of Associazione Donatori Sangue samples: these are indicated as 12CD, 28CD and 43CD.

The bioinformatic analysis revealed that there are not common haplotypes in PD patients *versus* HC and in none of the genotype subcategories.
Interestingly, although no CNVs were found in Nprl3 gene locus, the analysis revealed several genes subjected to copy number variation. On Table 7, the relative confidence score is indicated for every genomic location detected (genomic assembly hg38).
Further analysis and validation should be done.

Sample	Genomic location (hg38)	Copy number	Gene	Confidence score
10HC	chr5:113327484-113330738	1	MCC	14.672
10HC	chr5:139931629-139931653	1	NRG2	41.034
10HC	chr13:38970825-38992446	1	STOML3	21.842
10HC	chr20:14702391-14732585	1	MACROD2	30.908
10HC	chr2:79516139-79709354	3	CTNNA2	160.845
10HC	chr2:79516139-79709354	3	MIR4264	160.845
10HC	chr13:20425867-20436598	3	CRYL1	43.006
10HC	chr13:20425867-20436598	3	MIR4499	43.006

10HC	chr14:20404614-20404736	3	TEP1	16.048
10HC	chr16:28615243-28620752	3	SULT1A1	28.090
10HC	chr17:77381431-77383748	3	SEPT9	17.582
10PD	chr1:74649119-74649321	1	ERICH3	16.074
10PD	chr1:85029067-85029489	1	MCOLN3	21.550
10PD	chr1:91843626-91845778	1	TGFBR3	18.709
10PD	chr2:55544485-55563939	1	SMEK2	17.546
10PD	chr2:55544485-55563939	1	CFAP36	17.546
10PD	chr3:75428675-75605157	1	FAM86DP	20.471
10PD	chr3:121591417-121605219	1	FBXO40	13.749
10PD	chr4:70918817-70921282	1	MOB1B	18.832
10PD	chr4:102792983-102816505	1	UBE2D3	21.241
10PD	chr4:102792983-102816505	1	LOC102723704	21.241
10PD	chr5:19543155-19548361	1	CDH18	17.917
10PD	chr5:37138835-37140209	1	C5orf42	16.326
10PD	chr5:60179360-60186846	1	PDE4D	13.831
10PD	chr5:139931629-139931740	1	NRG2	66.498
10PD	chr6:56917508-56918076	1	DST	16.386
10PD	chr6:78972930-79029367	1	PHIP	22.046
10PD	chr6:144838041-144843858	1	UTRN	18.631
10PD	chr7:3611608-3615700	1	SDK1	15.299
10PD	chr7:18370302-18385624	1	HDAC9	14.877
10PD	chr7:23728916-23729075	1	STK31	15.891
10PD	chr7:48316076-48318680	1	ABCA13	30.944
10PD	chr7:112614215-112631636	1	LOC100996249	19.281
10PD	chr7:112614215-112631636	1	LOC101928012	19.281
10PD	chr9:17462978-17470126	1	CNTLN	13.975
10PD	chr9:105766479-105767091	1	TMEM38B	16.474
10PD	chr10:76361320-76373904	1	C10orf11	9.982
10PD	chr10:91470828-91483101	1	HECTD2	20.172
10PD	chr10:91470828-91483101	1	HECTD2-AS1	20.172
10PD	chr10:91497409-91498808	1	HECTD2	22.071
10PD	chr10:91497409-91498808	1	HECTD2-AS1	22.071
10PD	chr11:32635111-32637615	1	CCDC73	25.538
10PD	chr11:92887296-92887445	1	FAT3	20.323
10PD	chr12:21200019-21207479	1	SLCO1B1	25.176
10PD	chr12:29648262-29648367	1	TMTC1	21.187
10PD	chr12:40076736-40085906	1	SLC2A13	10.556
10PD	chr12:88490666-88524191	1	KITLG	31.882
10PD	chr13:93367328-93384397	1	GPC6	17.221
10PD	chr19:12541214-12542685	1	ZNF564	15.910
10PD	chr19:20626179-20728777	1	ZNF626	46.288
10PD	chr19:38159530-38160276	1	SIPA1L3	15.046
10PD	chr20:32996559-33001677	1	SUN5	12.689
10PD	chr20:58440630-58461828	1	VAPB	32.618
10PD	chr20:58440630-58461828	1	APCDD1L	32.618
10PD	chr7:17959134-18062796	3	PRPS1L1	126.170
11PD	chr3:37979882-37986249	1	CTDSPL	16.171
11PD	chr5:70305696-70307464	1	GUSBP3	20.113
11PD	chr5:97048466-97099320	1	LIX1	58.768
11PD	chr9:89031511-89035754	3	SHC3	33.825
12CD	chr5:151516652-151518156	0	FAT2	19.515
12CD	chr7:91033074-91038222	1	CDK14	17.716

12CD	chr8:115641238-115642408	1	TRPS1	12.481
12CD	chr10:47543322-47568296	3	BMS1P6	20.984
12CD	chr10:47646751-47703869	3	FAM35DP	52.489
12CD	chr19:43268069-43344686	3	PSG9	34.874
13PD	chr5:151514956-151518156	0	FAT2	12.747
13PD	chr3:89402447-89417171	1	EPHA3	24.117
13PD	chr5:139931633-139931658	1	NRG2	49.735
13PD	chr12:40534583-40538858	1	MUC19	17.207
13PD	chr19:41362173-41365095	1	TMEM91	24.060
13PD	chr19:41362173-41365095	1	B9D2	24.060
13PD	chr15:24619900-24728780	3	NPAP1	27.391
13PD	chr17:44169808-44248497	3	ASB16	53.405
13PD	chr17:44169808-44248497	3	ATXN7L3	53.405
13PD	chr17:44169808-44248497	3	UBTF	53.405
13PD	chr17:44169808-44248497	3	TMUB2	53.405
13PD	chr17:44169808-44248497	3	MIR6782	53.405
13PD	chr17:44169808-44248497	3	ASB16-AS1	53.405
13PD	chr17:44169808-44248497	3	SLC4A1	53.405
13PD	chr22:18969404-18972450	3	DGCR5	17.872
14HC	chr19:20626179-20715228	1	ZNF626	37.743
14HC	chr4:186071932-186134204	3	TLR3	54.038
14HC	chr9:135951231-135955672	3	UBAC1	14.142
14HC	chr21:14613203-14670124	3	LOC388813	10.898
14PD	chr12:40875351-40875963	0	CNTN1	21.682
14PD	chr5:139931607-139931653	1	NRG2	44.833
14PD	chr5:112915301-112947050	3	REEP5	105.297
14PD	chr6:78996260-79029367	3	PHIP	14.566
14PD	chr6:89157027-89162999	3	PM20D2	6.687
15HC	chr1:213009337-213011812	1	ANGEL2	16.549
15HC	chr19:43328006-43372386	1	PRG1	12.743
15HC	chr19:43328006-43372386	1	CD177	12.743
15HC	chr19:43401936-43539189	1	ETHE1	44.754
15HC	chr19:43401936-43539189	1	TEX101	44.754
15HC	chr19:43401936-43539189	1	PHLDB3	44.754
15HC	chr19:43401936-43539189	1	LYPD3	44.754
15HC	chr19:43401936-43539189	1	ZNF575	44.754
15HC	chr5:139931633-139931650	3	NRG2	34.058
15HC	chr10:47543322-47662831	3	BMS1P6	83.480
15HC	chr10:47543322-47662831	3	CTSLP2	83.480
15HC	chr10:47684765-47703869	3	FAM35DP	35.583
18PD	chr5:140232346-140238124	1	CYSTM1	20.919
18PD	chr6:78996260-79029367	1	PHIP	14.152
18PD	chr9:22765106-22819064	1	LINC01239	68.718
18PD	chr10:82879719-82890180	1	NRG3	20.581
18PD	chr2:38963211-38970130	3	ARHGEF33	21.544
18PD	chr3:155481097-155483319	3	PLCH1	28.305
18PD	chr5:139931607-139931740	3	NRG2	84.712
19PD	chr4:157305789-157306780	1	GRIA2	19.050
19PD	chr6:69687698-69690567	1	LMBRD1	13.773
19PD	chr14:35703521-35717282	1	RALGAPA1	27.761
19PD	chr15:55397388-55550755	1	PYGO1	122.328
19PD	chr15:55397388-55550755	1	DYX1C1-CCPG1	122.328
19PD	chr15:55397388-55550755	1	DYX1C1	122.328

19PD	chr15:55397388-55550755	1	CCPG1	122.328
19PD	chr15:55397388-55550755	1	C15orf65	122.328
19PD	chr5:139931633-139931740	3	NRG2	80.313
19PD	chr19:53521083-53539860	3	ZNF331	36.412
28CD	chr2:4651792-4653622	1	LINC01249	21.012
28CD	chr5:140229786-140238124	1	CYSTM1	37.270
28CD	chr5:177391049-177396392	1	SLC34A1	14.681
28CD	chr9:582371-594579	1	KANK1	27.552
28CD	chr16:28615243-28620752	1	SULT1A1	29.801
28CD	chr19:41362173-41365095	1	TMEM91	22.323
28CD	chr19:41362173-41365095	1	B9D2	22.323
28CD	chr6:78996260-79029367	3	PHIP	17.498
28CD	chr10:26822825-26873049	3	ABI1	45.732
2HC	chr4:157305789-157306780	0	GRIA2	12.835
2HC	chr5:139931633-139931661	1	NRG2	45.618
2HC	chr5:177391049-177396392	1	SLC34A1	14.791
2HC	chr8:7308659-7308718	1	FAM66B	9.637
2HC	chr15:22382531-22468503	1	GOLGA8CP	16.296
2HC	chr15:22382531-22468503	1	GOLGA8EP	16.296
2HC	chr15:34929955-34963958	1	AQR	48.099
2HC	chr3:197800244-197817520	3	LRCH3	23.279
2HC	chr10:42436301-42509794	3	CCNYL2	48.194
2HC	chr10:42436301-42509794	3	LINC00839	48.194
2HC	chr16:23072833-23078389	3	USP31	34.143
2HC	chr16:23098422-23113762	3	USP31	17.293
43CD	chr1:12919838-12920053	0	PRAMEF7	42.778
43CD	chr11:86304402-86305758	0	C11orf73	17.768
43CD	chr1:12907683-12919111	1	PRAMEF7	10.054
43CD	chr5:139931607-139931661	1	NRG2	48.429
43CD	chr8:115641238-115642408	1	TRPS1	13.669
43CD	chr10:46968630-47000252	1	PTPN20B	54.401
43CD	chr19:20626179-20647550	1	ZNF626	12.953
43CD	chr17:44177103-44189067	3	ASB16	13.762
43CD	chr17:44177103-44189067	3	TMUB2	13.762
43CD	chr17:44177103-44189067	3	ASB16-AS1	13.762
4PD	chr1:213002088-213011812	1	ANGEL2	23.198
4PD	chr5:70305696-70308271	1	GUSBP3	20.402
4PD	chr17:67074697-67081830	1	HELZ	42.878
4PD	chr19:20626179-20701612	1	ZNF626	34.735
4PD	chr5:139931633-139931740	3	NRG2	58.503
4PD	chr10:68418846-68429563	3	DNA2	34.027
4PD	chr16:28618318-28620752	3	SULT1A1	13.991
5HC	chr6:78972930-79029367	0	PHIP	36.350
5HC	chr3:75428675-75605157	1	FAM86DP	12.168
5HC	chr3:155481097-155487290	3	PLCH1	40.465
5HC	chr5:139931607-139931691	3	NRG2	77.170
5HC	chr18:63201728-63205702	3	BCL2	9.505
5PD	chr5:151514956-151518156	0	FAT2	21.616
5PD	chr1:174797867-174800121	1	RABGAP1L	16.765
5PD	chr12:31390640-31394235	3	DENND5B	18.478
5PD	chr15:101695029-101717607	3	TARSL2	42.299
5PD	chr16:6011798-6036851	3	RBFOX1	61.325
5PD	chr16:6044487-6046935	3	RBFOX1	27.379

5PD	chr16:6098295-6285091	3	RBFOX1	376.524

Table 7. CNV bioinformatic analysis.
The genomic coordinates of the CNV (genomic assembly hg38), the number of copies, the included gene and the confidence score are indicated.

Tag-containing transcript identification and validation

Identification of tag-containing transcripts

To identify the tag-containing transcripts I took advantage of 3'Rapid Amplification of cDNA Ends (RACE) applying this technique to the PD sample that presented the highest amount of tags on the nanoCAGE assay (PD13, Figure19).

PD13 RNA was retrotranscribed into cDNA modified with the GeneRacer 3' adaptor using poly-T primer and the first-step RT-PCR was carried out using the GeneRacer 3' Primer and a forward primer corresponding to the tag sequence (Figure 23A). Nested RT-PCR, performed using the GeneRacer 3' Nested Primer and the same forward primer, allowed amplification of a major product (about 2000 bp) (Figure 23E).

Subcloning and sequencing of this fragment revealed a new Nprl3 isoform with the TSS in the third intron of Nprl3 gene (Figure 23B).

I performed the 3'RACE assay several times and along this major product of amplification (that was always present in the PCRs) I have been able to subclone and sequence other minor products that could have an important biological meaning.

I identified four Nprl3-Hemoglobin A1 chimeric transcripts that I called transalpha 1-4. The breakpoints of these transcripts involve the intronic minisatellite sequence (2-4 repeats are transcribed) and the third exon of hemoglobin 1A (HBA1). Only in one case (transalpha4) I could observe a more complex transplicing composed by minisatellite intronic sequence spliced to the subsequent two exons of Nprl3 interrupted by part of the second and all the third exon of HBA1 (Figure 24).

As suggested by Unneberg and Claverie, 2007, chimeric transcripts may be the result of close encounters of active genes and could represent functional products or "noise" in the transcription process. Nprl3 gene lies in the distal region of the short arm of chromosome 16, which has been largely studied due to the presence of alpha globin gene cluster (Vyas et al., 1995). In particular HS-40, the major distal enhancer

of alpha globin cluster lying in the fourth intron of Nprl3, directly interacting with alpha gene promoters via chromatin looping, as proposed by Li, 2006 and demonstrated by Vernimmen et al., 2009.

This physical interaction may explain the chimeric transcripts formation but further studies will be necessary to unveil their functional role. Unfortunately, I could not be able to validate these chimeric transcripts due to the high abundance of HBA1 transcript in blood RNA samples.

I decided to repeat the 3'RACE experiment to confirm the existence of the new Nprl3 isoform. To validate the TSS located in the minisatellite I performed 5'RACE. To this purpose, I choose another high tag-expressing PD blood sample (PD11, Figure 19).

After the ligation of a synthetic RNA oligonucleotide to the 5' end of decapped RNA and retrotrascription into cDNA modified with the GeneRacer 3' adaptor using poly-T primer, in the 3'RACE experiment the first-step and the nested RT-PCR have been carried out in the same conditions as for PD13.

In the 5'RACE experiment first-step RT-PCR was carried out using the GeneRacer 5' Primer and a reverse primer located to the fourth exon of Nprl3 gene. 5'RACE nested RT-PCR was performed using the GeneRacer 5' Nested Primer and a nested reverse primer spanning the fourth Nprl3 exon and the transcribed intronic sequence found with 3'RACE (Figure 23C).

The 3'RACE gel run revealed a 2000 bp band and by subcloning and sequencing it I could confirm the result obtained with PD13 sample (data not shown). Moreover the 5'RACE gel run revealed the amplification of a major product (about 200 bp) (Figure 23F) and the subcloning and sequencing of this fragment confirmed that the TSS of this new Nprl3 isoform (TagNprl3) is located in the minisatellite sequence of Nprl3 third intron (Figure 23D).

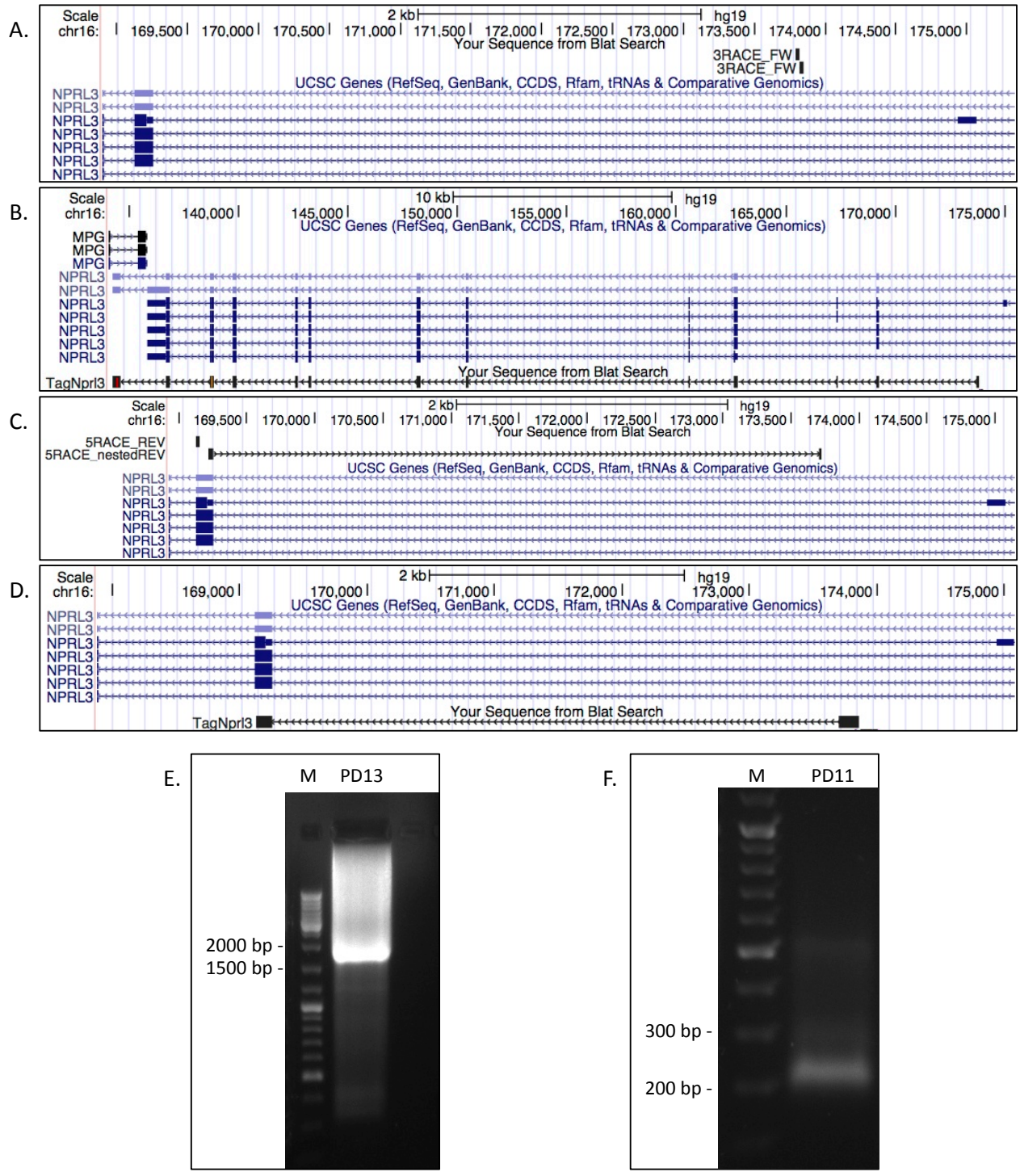

Figure 23. Identification of the tag containing transcripts by 3' RACE-PCR analysis and validation by 5' RACE-PCR analysis.
A) UCSC Genome Browser representation of the 3' RACE_FW primer (double mapping on the minisatellite sequence).
B) Sequence derived from 3'RACE product visualized with the USCS Genome Browser. 3'RACE validated the transcriptional activity of the minisatellite located in Nprl3 third intron.
C) UCSC Genome Browser representation of the 5' RACE_REV and 5'RACE_nestedREV primers.
D) Sequence derived from 5'RACE product visualized with the USCS Genome Browser. 5'RACE validated the transcription start site located in the intronic minisatellite.
E) 5'RACE performed on first-strand cDNA prepared from total RNA from PD11 whole blood. PCR products were analyzed on a 1% agarose gel with EtBr staining.
F) 3'RACE performed on first-strand cDNA prepared from total RNA from PD13 whole blood.

PCR products were analyzed on a 1% agarose gel with EtBr staining.

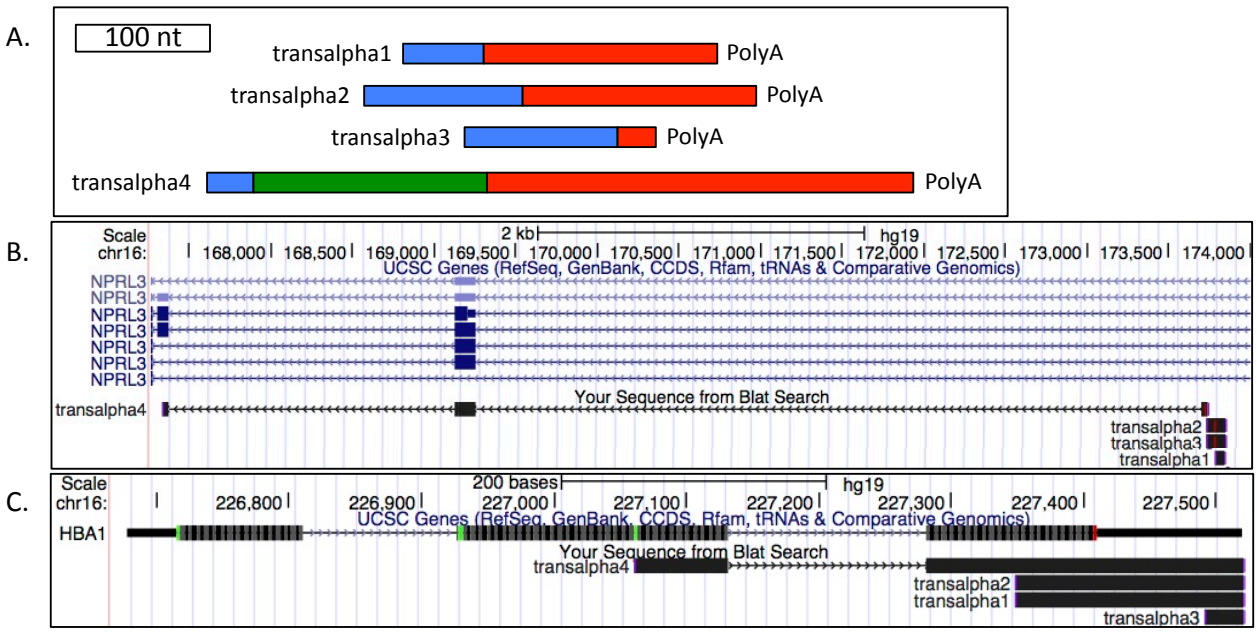

Figure 24. Identification of chimeric products by 3' RACE-PCR analysis.
A) Schematic representation of chimeric products found with 3'RACE analysis. The white block indicates the scale. In blue, Nprl3 intronic part; in green, Nprl3 exonic part; in red, Hba1 exonic part.
B) UCSC Genome Browser representation of 5' ends of the chimeric transcripts on Nprl3 gene.
C) UCSC Genome Browser representation of 3' ends of the chimeric transcripts on Hba1 gene.

I then focused my attention on the TSS to identify the minisatellite repeat that contains the core promoter. I thus discovered that the +1 nucleotide of the transcript is a cytosine located on repeat number 11 (Figure 25).

The most commonly occurring core promoter motif in mammalian cells is the Inr initiator which consensus sequence is YYA+1NWYY in humans (Javahery et al., 1994). However, the sequence for TagNprl3 transcription is TC+1TTTCT which strongly resemble a TCT motif (YC+1TYTYY), a key component of an RNA polymerase II system that is directed toward the expression of ribosomal protein genes as well as other genes encoding factors involved in protein synthesis (Parry et al., 2010).

Despite the high sequence homology between the different repeats, the consensus TC+1TTTCT is present only in the eleventh repeat and could explain the specificity of transcription start, however it is present both in 16 and 13 repeats alleles so it could not explain the differential expression between 16/13 and 16/16 individuals.

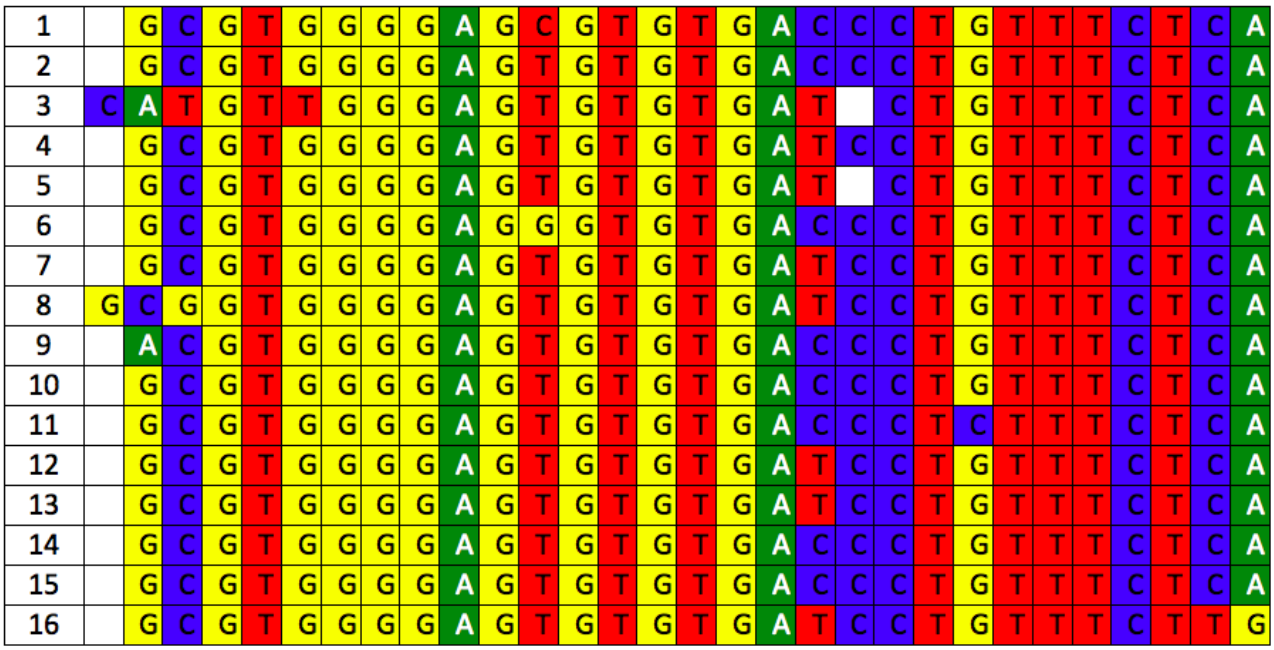

Figure 25. Schematic representation of minisatellite 16 repeats.
The transcription starts on the cytosine at position 22 in the eleventh repeat. The TC+1TTTCT consensus present only in this repeat strongly resemble a TCT motif (YC+1TYTYY), a key component of an RNA polymerase II system that is directed toward the expression of ribosomal protein genes as well as other genes encoding factors involved in protein synthesis (Parry et al., 2010). The 13-repeats allele lacks lines 7-8-9 and so TC+1TTTCT consensus is conserved.

TagNprl3 transcript validation

To validate the TagNprl3 transcript identified with RACE, I designed a reverse primer in the 3'UTR region. I performed RT-PCR by using this primer in combination with the same forward primer of the 3'RACE assay, which corresponds to the nanoCAGE tag sequence (Figure 26A). Together with PD13 16/13, I decided to test another heterozygous high-expressing tag sample (PD2, 16/13, Figure 19) and a homozygous low-expressing tag sample (HC14, 16/16, Figure 19).

As expected, the amplicon is around 2000 bp and by sequencing I confirmed the identity of the TagNprl3 sequence. Together with validating TagNprl3 transcript, I also proved that TagNprl3 is indeed differentially expressed between high- and low-expressing tag samples, as expected (Figure 26B).

On the contrary, by performing an RT-PCR with Nprl3 isoform 1 primers, its expression is quite comparable among the three samples (Figure 26C). This pair of primers will be used in the Real-Time TaqMan PCR.

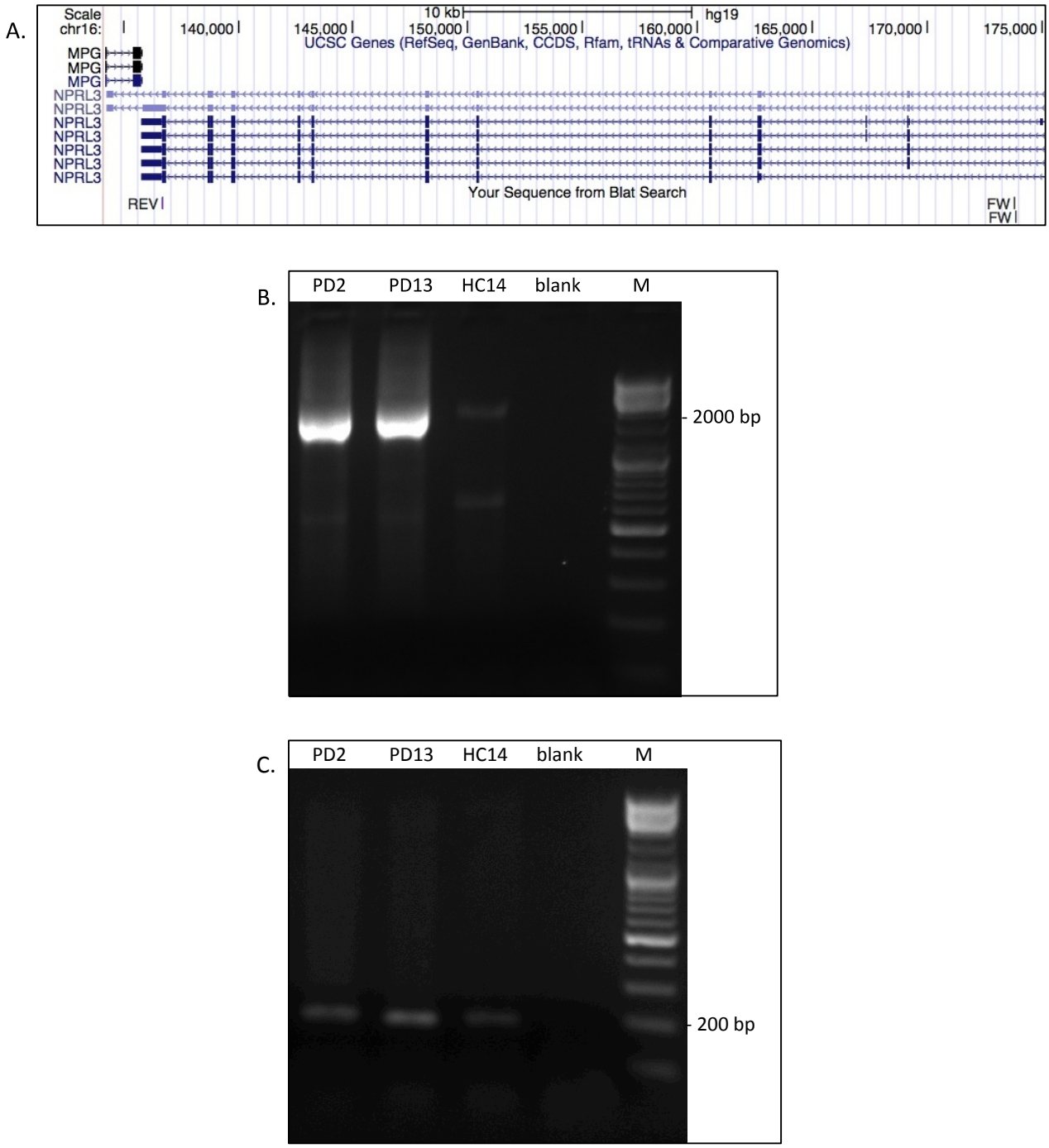

Figure 26. Validation of TagNprl3 transcript.
A) UCSC Genome Browser representation of the FW primer (double mapping on the minisatellite sequence) and the REV primer used for TagNprl3 validation.
B) TagNprl3 RT-PCR performed on first-strand cDNA prepared from total RNA from PD2, PD13 and HC14 whole blood. PCR products were analyzed on a 1% agarose gel with EtBr staining. The main 2000 bp product corresponds to TagNprl3 and its expression is barely detectable in HC14 individual (16/16, low-tag expressing) wherease is intense in PD2 and PD13 (both 16/13, high-tag expressing).
C) Nprl3 RT-PCR performed on the same samples (PD2, PD13 and HC14). PCR products were analyzed on a 1% agarose gel with EtBr staining. The 200 bp product corresponds to Nprl3 and its expression is comparable among the three individuals.

TagNprl3 expression is linked to minisatellite genotype

To assess if high TagNprl3 expression correlates with high nanoCAGE tag expression I selected 19 RNA samples that have been used for the nanoCAGE analysis: 6 PD high-expressing tag, 6 PD low-expressing tag, 6 HC low-expressing tag and the HC with the highest number of tags.

I carried out Real-Time TaqMan RT-PCR with the assays for TagNprl3 and Nprl3 mRNAs. I took advantage of PGK1 as invariant gene for normalization.

Nprl3 expression levels are quite variable among different samples but show no correlation with nanoCAGE tag expression or disease condition (Figure 27A).

On the other hand, TagNprl3 high expression is strongly correlated with high nanoCAGE tag expression while it is low when tag was almost undetectable. Unexpectedly, HC15 expresses high levels of TagNprl3 (Figure 27B).

It is then important to clarify whether TagNprl3 high expression correlates with the 16/13 minisatellite genotype or with PD.

To address this crucial question I need to collect a larger number of HCs with 16/13 genotype.

To this purpose, I created an independent set of samples (PD from "Clinica Neurologica di Trieste" and "Telethon Biobank", HC from "Associazione Donatori Sangue Trieste") that have been genotyped for the minisatellite sequence in order to collect a sufficient amount of heterozygous individuals. This set of samples is composed by 6 PD 16/16, 4 PD 16/13, 6 HC 16/16 and 4 HC 16/13 (details on Table 8).

I then performed a Real-Time TaqMan RT-PCR in the same conditions as for the validation set and with the same assays (TagNprl3, Nprl3 and PGK1 for normalization).

Nprl3 expression is variable among individuals but I could not find any correlation between its expression and genotype or disease state (Figure 27C).

TagNprl3 expression is clearly linked to 16/13 minisatellite genotype being highly induced in these individuals compared to homozygous however, I could not find any correlation with PD (Figure 27D).

In summary, TagNprl3 is a new Nprl3 isoform with a new TSS in its third intron and its expression is highly induced by the genotype condition of the minisatellite lying in this intron.

Experiment	Minisatellite Genotype	Sample Name
Validation	16/16	2C
		5C
		7C
		8C
		10C
		14C
	16/13	15C
	16/16	1PD
		6PD
		8PD
		9PD
		10PD
		19PD
	16/13	2PD
		4PD
		7PD
		11PD
		13PD
		18PD
Independent	16/16	PD_C2616
		PD_M2371
		PD_Z0411
		PD_S1613
		PD_T0770
		PD_C2766
	16/13	PD_025
		PD_022
		PD_Z0397
		PD_G1593
	16/16	C_49AB
		C_51PR
		C_52FC
		C_54MK
		C_47IC
		C_48MK
	16/13	**C_12CD**
		C_28CD
		C_43SB
		C_43CS

Table 8. List of Real Time TaqMan PCR blood samples and their minisatellite genotype. The two samples in bold have been used for the fractionation experiment.

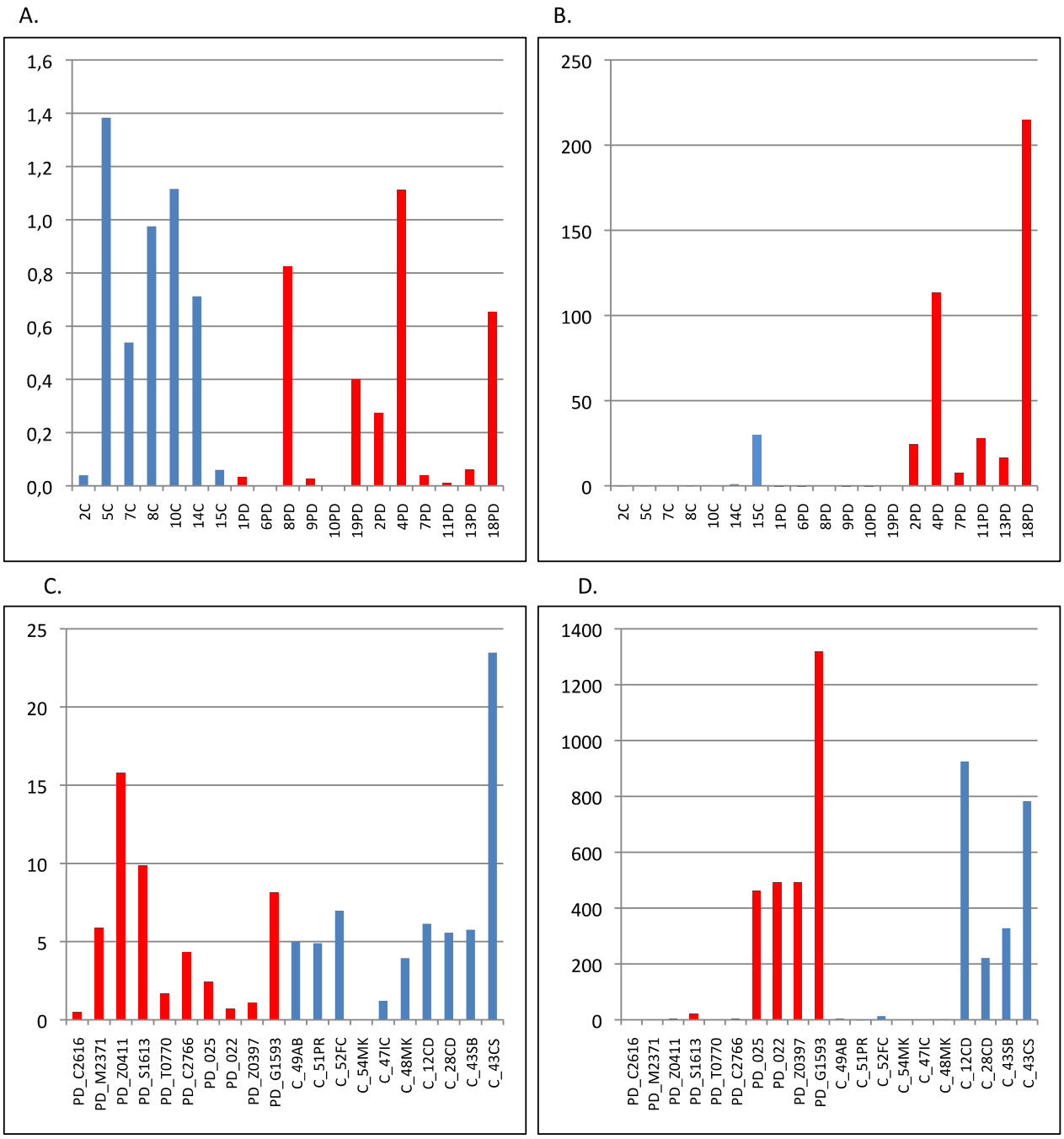

Figure 27. TagNprl3 upregulation in blood is linked to 16/13 minisatellite genotype.
Real-Time TaqMan PCR normalized with PGK1 assay. Fold change is indicated on the y-axis. PD are indicated in red and HC in blue.
A) Nprl3 assay on selected nanoCAGE samples. Canonical Nprl3 isoform expression is variable among individuals but it is independent from minisatellite genotype or healthy/pathological state.
B) TagNprl3 assay on selected nanoCAGE samples reveals that its high expression is linked to 16/13 minisatellite genotype.
C) Nprl3 assay on an independent sample set (PD and HC, 16/13 and 16/16, see Table 8). I could confirm that Nprl3 expression is independent from minisatellite genotype or healthy/pathological state.
D) TagNprl3 assay on an independent sample set (PD and HC, 16/13 and 16/16, see Table 2) confirmed that all 16/13 individuals (both PD and HC) express high levels of TagNprl3.

Biological function of TagNprl3

TagNprl3 encodes for a nucleo-cytoplasmic protein

Nprl3 gene has 5 annotated different alternative splicing transcripts. In particular, isoforms 2-5 are predicted to translate a shorter Nprl3 protein, starting from an alternative AUG. Until now, only Nprl3 isoform 1 (NM_001077350) is known to encode for a protein of 64 kDa, while truncated forms were undetectable in several cell types (Lunardi et al., 2009). The coding sequence of TagNprl3 corresponds to the coding sequence of Nprl3 isoform 3 (NM_001243247) but their UTRs are different. Nprl3 isoform 3 is an N-terminal truncated (Δ1-78 aminoacids) Nprl3 protein.

To investigate the function of TagNprl3 I cloned the full-length sequence into pcDNA3.1- expression vector.

First, I wanted to investigate whether the TagNprl3 transcript encodes for the N-terminal-truncated Nprl3 protein. To this purpose I transiently transfected HEK 293T cells with TagNprl3 using Nprl3 as control. By taking advantage of an antibody against Nprl3 (Abcam) I observed that TagNprl3 encodes for a 60 kDa protein while full-length Nprl3 protein is of the expected size (64 kDa, lane 2) (Figure 28A).

Then I studied TagNprl3 subcellular localization with an immunofluorescence assay on transiently transfected HEK 293T cells and I observed a similar nucleo-cytoplasmic distribution for Nprl3 and TagNprl3 (Figure 28B).

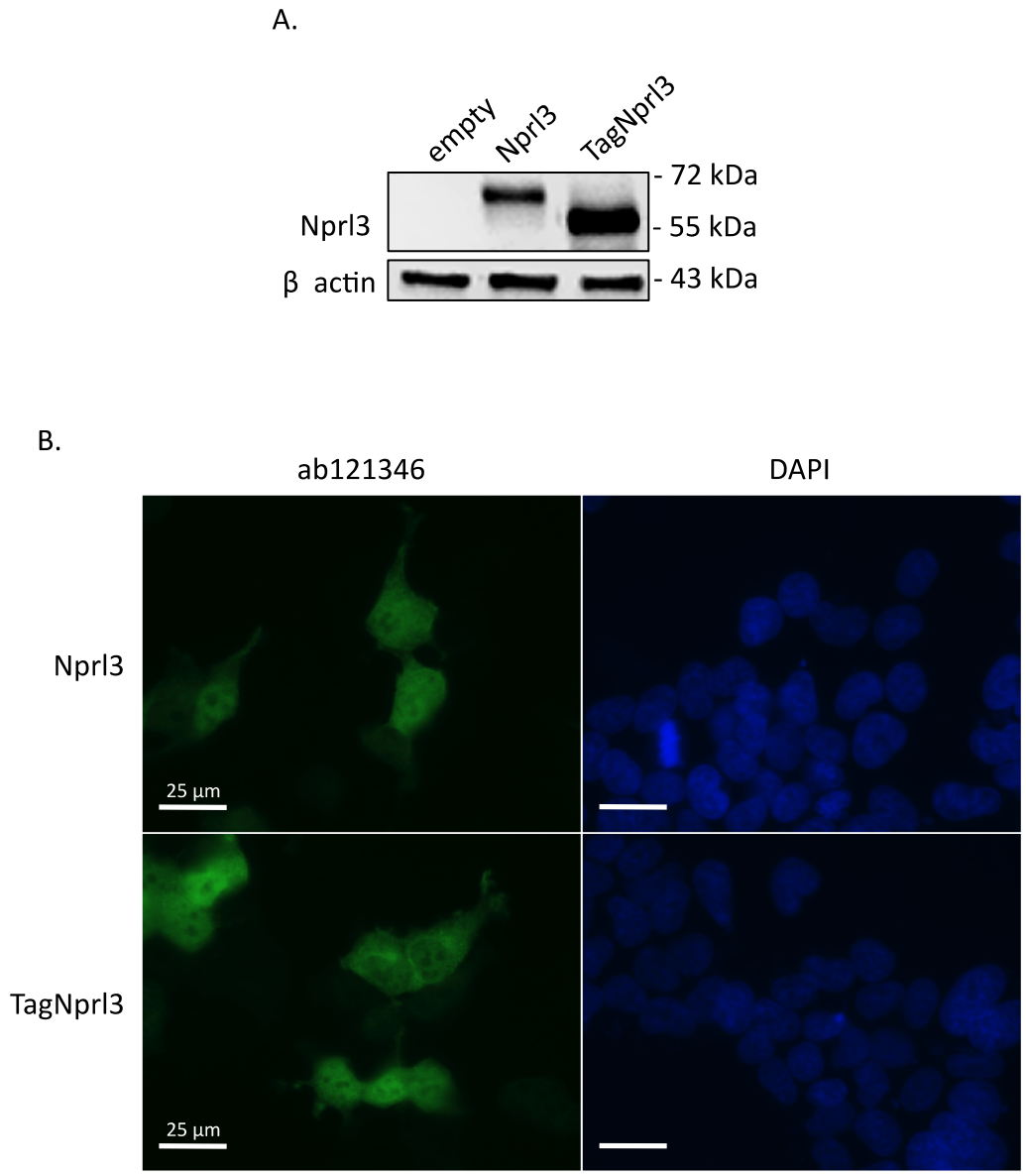

Figure 28. Analysis of the exogenous expression of TagNprl3 reveals a 60 kDa protein that has a nucleo-cytoplasmic localization.
A) The full-length transcript of TagNprl3 was cloned into an expression vector and transiently transfected in HEK 293T cells. Taking advantage of an antibody against Nprl3 (Abcam) I could demonstrate that TagNprl3 translation give rise to a protein product of about 60 kDa. HEK cells transiently transfected with canonical Nprl3 have been used to appreciate the molecular weight difference. β actin was used as a loading control.
B) Immunofluorescence analysis on HEK 293T cells transiently transfected with Nprl3 or TagNprl3 and stained with anti-Nprl3 antibody reveals that the two proteins have a similar nucleo-cytoplasmic distribution.

TagNprl3 binds Nprl2

It has been demonstrated that Nprl3 forms an heterodimer with its partner Nprl2 and, together with DEPCDC5, they form a protein complex called GATOR1 in mammals, which in turns is involved in many cellular processes such as stimulation of protein

synthesis, growth, metabolism and the inhibition of autophagy (Bar-Peled et al., 2013).

To assess if TagNprl3-encoded protein (that corresponds to Nprl3 isoform 3) could interact with Nprl2 a co-immunoprecipitation assay has been set up by co-transfecting HEK cells with Nprl2_MYC and TagNprl3. As control I used Nprl3 isoform 1 and Nprl3 isoform 2: the former is the canonical Nprl3 isoform involved in GATOR1 complex, the latter is a N-terminal truncated (Δ1-179 aminoacids) protein-coding isoform.

Our hypothesis is that Nprl3 isoform 2 should not co-immunoprecipitate because it lacks the N-terminal longin domain that should dimerize with the N-terminal longin domain of Nprl2, according to Levine et al., 2013.

As shown in Figure 29, Nprl2 unexpectedly interacts with all the three different Nprl3 isoforms meaning that there could be at least one other domain involved in the dimerization. Moreover, Nprl3 isoform 1 transcript encodes, to a lesser extent, for all the other isoforms by alternative translation initiation. Interestingly, Nprl2 showed a higher affinity for N-terminal truncated 60 kDa protein encoded by Nprl3 isoform 3, the protein encoded by TagNprl3 (Figure 29).

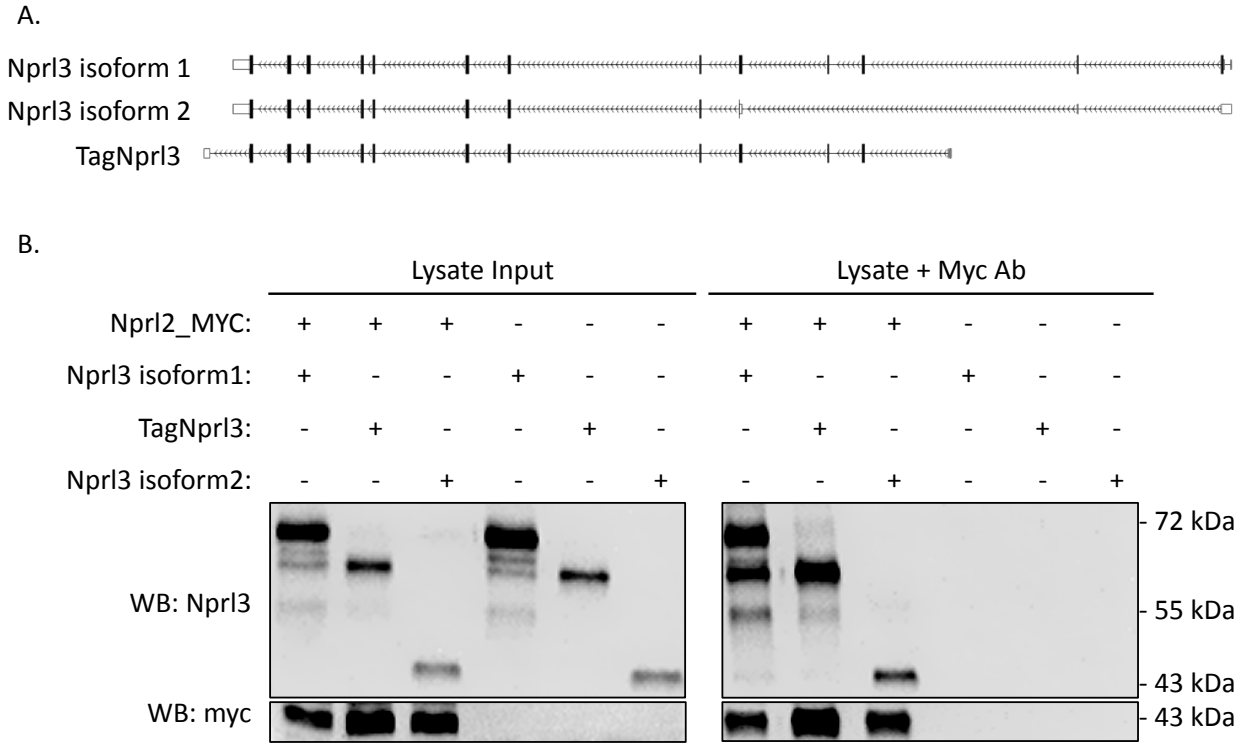

Figure 29. Co-immunoprecipitation of Nprl2 with Nprl3 isoforms.
A) Schematic representation of the three Nprl3 isoforms tested for Nprl2 interaction. The sense of transcription is from right to left.

B) Western blot showing the co-immunoprecipitation assay in HEK 293T cells transiently transfected with Nprl2 and one of the three Nprl3 isoforms. After 24 h from transfection, cell lysates have been incubated with Myc antibody (Cell Signaling) that recognize the C-terminal tag of Nprl2. Protein G Sepharose beads have been used to precipitate the protein complexes bound by Myc antibody that have been tested for Nprl3 isoforms presence by western blot. Surprisingly, Nprl2 co-immunoprecipitates with all the three Nprl3 isoforms.

TagNprl3 inhibits proliferation

By carrying out co-immunoprecipitation experiments, I observed that plates containing TagNprl3-transfected cells showed a marked decrease in cell counts.

It has been already demonstrated that Nprl3 inhibits proliferation (Lunardi et al., 2009) and I therefore took advantage of the BrdU assay to test whether TagNprl3 was exerting the same effect.

HEK 293T cells were transfected with plasmids expressing GFP, Nprl3 or TagNprl3 and after 48 hours assayed with BrdU. By immunofluorescence with anti-BrdU antibodies, I found that Nprl3- and TagNprl3-transfected cells showed a significative decrease in the rate of proliferation compared to controls: 40% of GFP positive cells were growing in control experiment while only about 15% of Nprl3-overexpressing cells and about 20% of TagNprl3 cells incorporated BrdU.

I thus confirmed Lunardi et al. data and proved that TagNprl3 has a similar function in inhibiting cell proliferation.

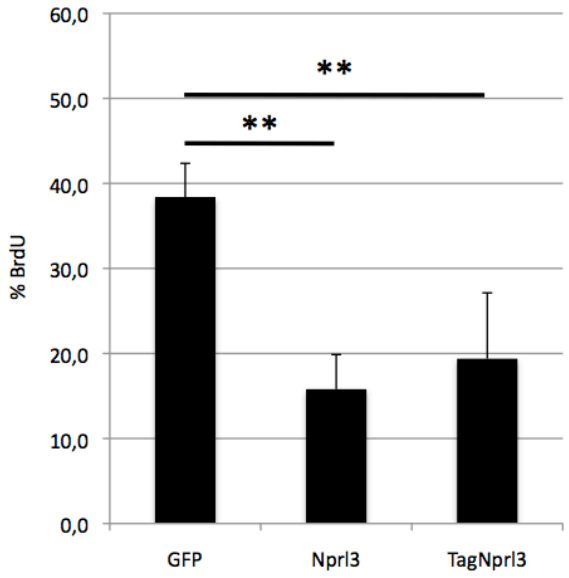

Figure 30. TagNprl3 retains the ability of the full-length protein to inhibit proliferation. Nprl3 and TagNprl3 inhibit BrdU incorporation. A vector expressing GFP was transfected as a control. The y axis indicates the % of BrdU positive cells in the transcfected population. Number of replica=5.

TagNprl3 overexpression has no effect on mTORC1

As part of GATOR1 complex, Nprl3 plays an important role in the negative regulation of mTORC1 and its expression alterations might influence downstream effectors. I therefore decided to analyze by western blot two different pathways that are under the control of mTORC1: protein synthesis, through the ribosomal protein S6 kinase (S6K) and 4E-binding protein 1 (4E-BP1) that are phosphorylated by active mTORC1, and autophagy, through microtubule-associated protein light chain 3 (LC3) that is converted in the active form if the process is stimulated.

As shown in Figure 31, mTORC1 stimulates protein synthesis through S6K stimulation and 4E-BP1 inhibition and at the same time inhibits autophagy. GATOR1 complex in turn inhibits mTORC1, as well as the common antitumoral drug Rapamycin that was used as a control of pathways alteration.

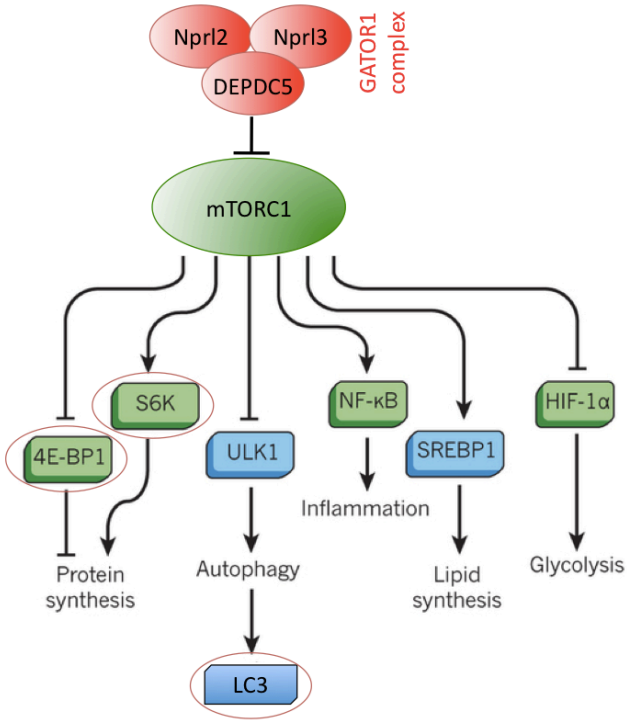

Figure 31. mTORC1 effectors. The involved molecular pathways are indicated. GATOR1 complex directly inhibits mTORC1 as well as Rapamycin drug (not indicated). The proteins that have been assayed by western blot are circled in red (adapted from Johnson et al., 2013).

As shown in Figure 32, the effect of Rapamycin inhibiting mTORC1 is evident in the loss of phosphorylation of both p70 S6K and 4EBP1. However this treatment is not sufficient to trigger autophagy as the conversion of soluble LC3-I (19 kDa) to active lipid bound LC3-II (17 kDa) is not visible.

Upon Nprl3 and TagNprl3 I could not observe alterations in the phosphorylation state of S6K nor 4EBP1, indicating that mTORC1 is not inhibited. On the other

hand, I could observe the appearance of a faint signal corresponding to active LC3-II in TagNprl3-transfected sample however, this effect must be further investigated.

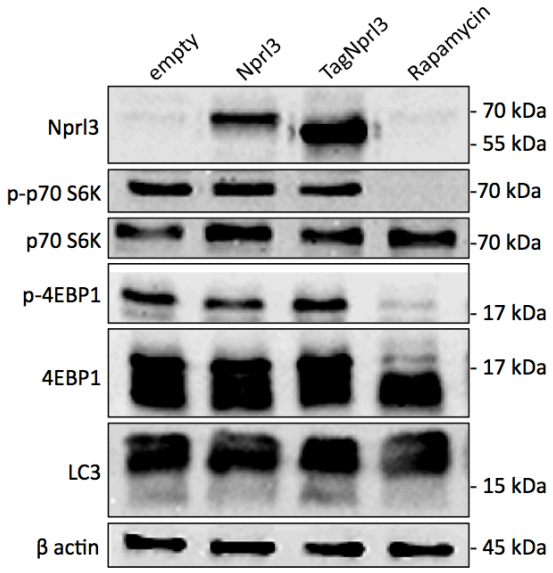

Figure 32. mTORC1 downstream effectors are not altered upon Nprl3 nor TagNprl3 overexpression. HEK 293T cells have been transiently transfected with Nprl3, TagNprl3 or with empty vector. As a control of mTORC1 inhibition, cells were treated for 1 hour with 2 µM Rapamycin. The western blot analysis of p-p70 S6K and p-4EBP1, that are protein synthesis markers downregulated upon mTORC1 inhibition, indicates that Nprl3 or TagNprl3 overexpression alone is not sufficient to trigger this effect. The autophagy marker LC3 indicated that autophagy is not stimulated but the appearance of active LC3-II is not detectable neither upon Rapamycin treatment. β-actin was used as a normalization control.

RBCs are the main source of TagNprl3 in blood

I wanted to assess what is the main source of TagNprl3 expression among the three different blood components: peripheral blood mononuclear cells (PBMC), red blood cells (RBC) and plasma.

I previously showed that TagNprl3 expression is not linked to PD but is driven by genetic heterozygosity of the minisatellite. Therefore I decided to take advantage of two 16/13 HCs part of the Real-Time TaqMan PCR validation set (C_43_CS and C_12_CD, see Table 8) to be compared to two 16/16 HC (MB and RC).

The blood has been collected into EDTA tubes and then the three blood fractions have been separated with Histopaque®-1077 reagent from SIGMA-ALDRICH. After RNA extraction, I prepared cDNA and performed a Real-Time TaqMan RT-PCR with the assays TagNprl3, Nprl3 and PGK1 for normalization.

Nprl3 and TagNprl3 expression is quite exclusive of RBC fraction and, in heterozygous individuals, it is highly induced. PBMC express low levels of Nprl3

while TagNprl3 was undetectable except for individual C_43_CS. However, for this sample I cannot exclude a contamination due to the high induction of TagNprl3 mRNA (Figure 33A, 33B). Both Nprl3 and TagNprl3 were undetectable in plasma (data not shown).

The main source of TagNprl3 are the RBCs and surprisingly, its expression revealed to be higher that the canonical Nprl3 isoform 1 in homozygous individuals.

TagNprl3 protein is upregulated in the RBCs of heterozygous individuals

I wanted to assess if TagNprl3 expression in RBCs gives rise to a protein product that is highly induced in heterozygous individuals. I thus prepared protein lysates from the RBC of MB 16/16 healthy individual, and from two 16/13 healthy individuals (C_43_CS and C_12_CD) that have been previously assayed for TagNprl3 RNA.

These protein lysates have been assayed by western blot: with anti-Nprl3 antibody I could detect a band of about 60 kDa in all the human samples. Moreover, this protein is highly induced in the two 16/13 individuals suggesting that TagNprl3 is translated *in vivo* and its expression is upregulated by heterozygous minisatellite genotypes (Figure 33).

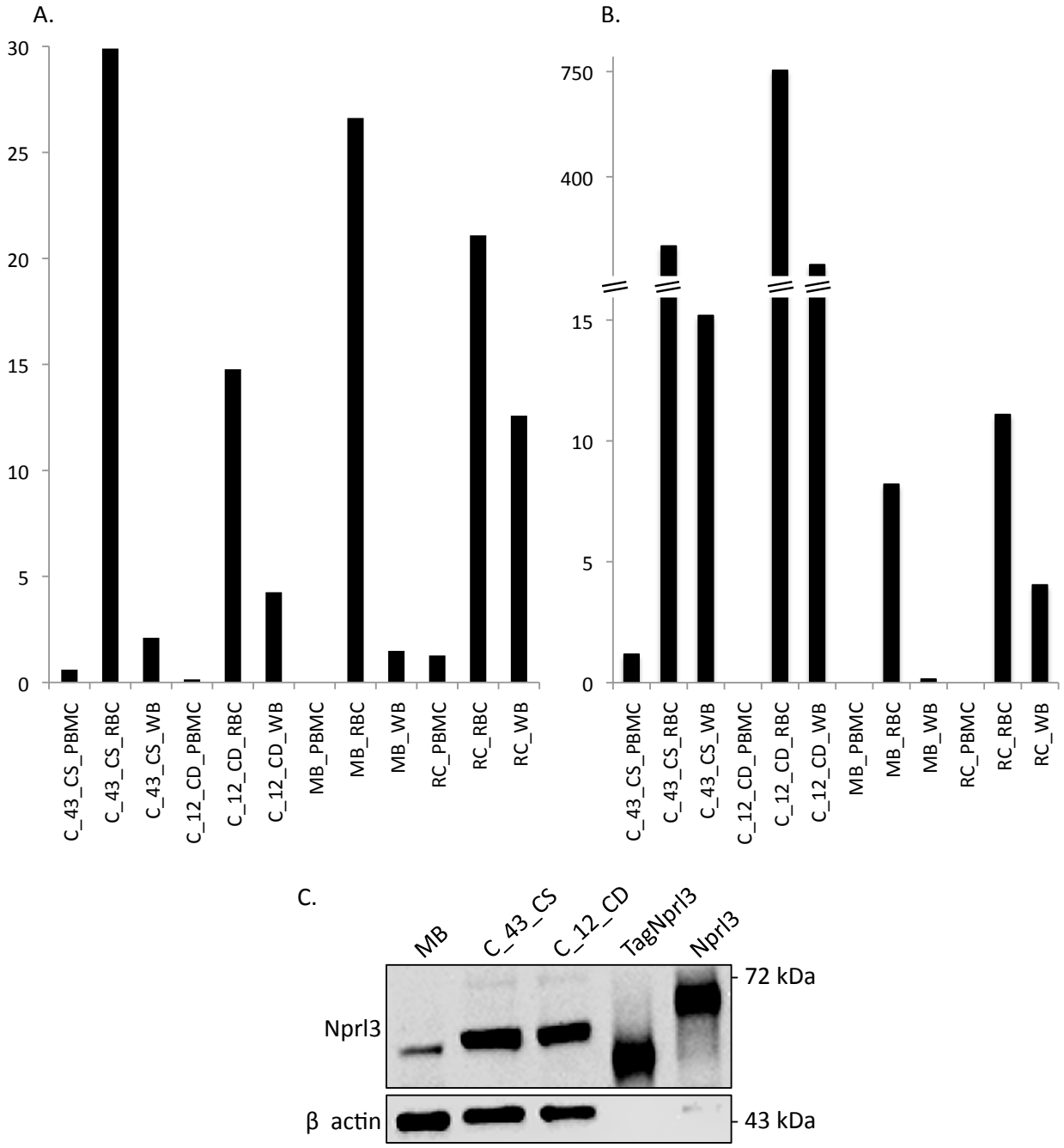

Figure 33. Red blood cells (RBC) are the main source of TagNprl3 expression in blood and 16/13 minisatellite genotype induces TagNprl3 transcript and protein upregulation.
For these experiments 4 HC have been chosen: 2 heterozygous (C_43_CS and C_12_CD) and 2 homozygous (MB and RC). Whole blood samples have been fractionated with Histopaque®-1077 reagent (SIGMA-ALDRICH) and the total RNAs have been extracted with TRIzol reagent (Invitrogen). After cDNA preparation I performed two Real-Time TaqMan assays:
A) Nprl3 assay shows that its expression is variable among the four individuals and ubiquitous in the cell fractions, although prevalent in RBC.
B) TagNprl3 assay shows that its expression seems to be exclusive of RBC (except from C_43_CS PBMC) and it is highly induced in the RBC of heterozygous individuals.
C) Western blot on RBC fraction protein lysate of a homozygous individual vs two heterozygous individuals shows that TagNprl3 protein exists *in vivo* and its induction by 16/13 genotype corresponds to a higher protein production. HEK cells transiently transfected with TagNprl3 or Nprl3 have been used as molecular weight control. Membranes have been incubated with anti-Nprl3 antibody (Abcam) and anti-β actin antibody has been used as normalizator.
PBMC: Peripheral Blood Mononuclear Cells; RBC: Red Blood Cells; WB: Whole Blood.

TagNprl3 expression in other tissues

I then asked if TagNprl3 expression could be detectable in other tissues. To this purpose I interrogated Zenbu database on human FANTOM5 (Functional Annotation of the Mammalian Genome, established by Dr. Hayashizaki and his colleagues in 2000, Riken, Japan).

Zenbu is a data integration, data analysis, and visualization system enhanced for RNAseq, ChipSeq, CAGE and other types of next-generation-sequence-tag (NGS) based data.

I interrogated the database at the genomic coordinates around TagNprl3 TSS: chr16:173854-173884 (hg19) and observed that the highest values of tag per million (tpm) are present in blood-derived samples.

On Table 10 I report all the cell types that have a tpm value in this genomic position. Although Nprl3 expression is detectable in all cell types and higher in blood tissue, TagNprl3 seems to be expressed exclusively in blood. However, validation is needed as well as comparison with analysis of 16/13 individuals.

cell type	**tpm**
Whole blood (ribopure), donor090309, donation2 : CNhs11671 ctss	2.92
Whole blood (ribopure), donor090309, donation1 : CNhs11675 ctss	2.57
Whole blood (ribopure), donor090612, donation2 : CNhs11673 ctss	2.39
Whole blood (ribopure), donor090612, donation1 : CNhs11672 ctss	1.69
Whole blood (ribopure), donor090309, donation3 : CNhs11948 ctss	1.16
blood, adult, pool1 : CNhs11761 ctss	1.12
mesenchymal stem cells (adipose derived), adipogenic induction, 00hr30min, biol rep2 : CNhs13426 ctss	0.67
CD34 cells differentiated to erythrocyte lineage, biol rep1 : CNhs13552 ctss	0.47
Myoblast differentiation to myotubes, day12, Duchenne Muscular Dystrophy donor3 : CNhs14613 ctss	0.41
hIPS +CCl2, biol rep1 : CNhs14217 ctss	0.34
Schwann Cells, donor3 : CNhs12621 ctss	0.31
Whole blood (ribopure), donor090612, donation3 : CNhs11949 ctss	0.27
COBL-a rinderpest(-C) infection, 06hr, biol rep3 : CNhs14437 ctss	0.26
Whole blood (ribopure), donor090325, donation2 : CNhs11076 ctss	0.26
MCF7 breast cancer cell line response to EGF1, 01hr40min, biol rep3 : CNhs12743 ctss	0.21
CD34 cells differentiated to erythrocyte lineage, biol rep2 : CNhs13553 ctss	0.18
SABiosciences XpressRef Human Universal Total RNA, pool1 : CNhs10610 ctss	0.16
Astrocyte - cerebral cortex, donor2 : CNhs11960 ctss	0.16
lung, fetal, donor1 : CNhs11680 ctss	0.16
Natural Killer Cells, donor3 : CNhs12001 ctss	0.11
spleen, fetal, pool1 : CNhs10651 ctss	0.09
Universal RNA - Human Normal Tissues Biochain, pool1 : CNhs10612 ctss	0.07
kidney, fetal, pool1 : CNhs10652 ctss	0.07

Table 10. FANTOM5 tag per million values for human samples at genomic location chr16:173854-173884 (hg19).

Discussion

PD is a neurodegenerative disorder that affects 1-2 % of all population above the age of 65. PD is a systemic, multifactorial disease caused by both environmental and genetic factors. The population aging together with the exposure to industrial and rural toxins seemed to be the main cause of its prevalence.

Unfortunately, at present there is no pharmacological treatment that can revert or slow neurodegeneration. Its diagnosis remains difficult since no standard tests are available. Patients are usually diagnosed at a late stage of PD, when motor symptoms appear and the loss of DA-producing cells in the SNc is about 60-70% (Dauer and Przedborski, 2003).

The strong need of an effective and early diagnosis has stimulated PD research towards a quest for novel biomarkers by gene expression analysis. In this context, since PD is a systemic disease, alterations associated to the pathology should be found outside the central nervous system.

One of the most promising tissues for biomarker discovery is blood since it is accessible and easy to collect with a fast and reproducible procedure. Furthermore, many works have interrogated the blood transcriptome of neurodegenerative disease patients, including PD, in search for a gene signature of the disease (some examples are Infante et al., 2015; Molochnikov et al., 2012; Scherzer et al., 2007; Zhang et al., 2014). Despite these attempts, a reproducible blood gene signature for PD is still lacking.

The laboratory of Professor Gustincich has recently carried out the largest study to date on gene expression analysis of the peripheral blood of sporadic drug naïve and *de novo* PD patients (Calligaris et al. 2015, in press). This work analyzed 40 sporadic PD patients and 20 HCs ("Discovery set") by taking advantage of the Affymetrix platform. Patients were at the onset of motor symptoms and before initiating any pharmacological treatment. By applying bioinformatics analysis, gene expression profiling was able to discriminate patients from HCs and identified 54 differentially expressed genes in blood. Together with neuronal apoptosis, lymphocyte activation and mitochondrial dysfunction, already found in previous analysis of PD blood and *post-mortem* brains, they unveiled transcriptome changes enriched in biological

terms related to epigenetic modifications including chromatin remodelling and methylation.

I then decided to take advantage of a new technology called nanoCAGE, developed by the collaboration between the laboratories of Professor Gustincich and Professor Carninci in RIKEN (Japan) (Plessy et al., 2010). Cap Analysis of Gene Expression (CAGE) is a technology based on the generation of short sequence tags from the 5' end of full-length cDNAs followed by high-throughput sequencing. When mapped to a reference genome, CAGE tags survey TSSs activity of specific promoters and measure expression levels on a massive scale (Carninci et al., 2006; Suzuki et al., 2009). With nanoCAGE, cDNA tags corresponding to the first 27 nucleotides of all transcripts are synthesized and identified by next generation sequencing. The number of sequencing reads for a defined tag is directly proportional to its RNA template, providing a measure of its expression level.

This technique gives an unbias description of the cellular transcriptome. It can work on tiny amounts of RNAs, and allows studying the expression of both polyA$^+$ and polyA$^-$ RNAs.

In summary, nanoCAGE is a powerful tool to obtain a snapshot of the blood transcriptome of PD patients, analyzing both annotated and previously unidentified transcripts.

We therefore assayed by nanoCAGE the blood transcriptome of 20 PD drug naïve de novo PD patients *versus* 20 HC age and sex matched. 10 million tags were sequenced for each sample.

From the bioinformatic analysis the most up-regulated nanoCAGE tag mapped to chr16:173660-174239 region and it was highly induced (50/100 fold) in 30% of PD patients.

As expected from the multi-mapping nature of the tag, it mapped to a repetitive DNA region: a minisatellite with a unit length of 29 nt repeated 16 times in the human reference genome *hg19*. This minisatellite is human-specific and it is exclusively located in this region.

While no studies are available to date on this specific minisatellite, it maps in a highly studied locus: the alpha-globin genes cluster that regulates α-like globin chains expression.

This cluster is located in the telomeric region of chromosome 16 (16p13.3) and it includes an embryonic gene (ζ), two pseudogenes (θ and αD) and the hemoglobin alpha genes (α1 and α2), surrounded by widely expressed genes.
The main regulatory regions are located about 25-65 kb upstream of the alpha-globin genes and they are called Multispecies Conserved Sequences (MCS), called MCS-R1 to -R4 (Higgs, 2013).
In particular, MCS-R3 is the nearest regulatory element to the minisatellite, from which the distance is about 3 kb, and they both lie in the third intron of the widely expressed gene Nprl3.
By amplifying the minisatellite genomic portion by PCR we discovered that individuals present different combinations of two alleles: one with 16-repeats and one with 13 of the 29-mer sequence unit.
We then analysed the identity of the minisatellite variants in the 40 individuals of the cohort of the nanoCAGE study. Interestingly, we found that high-tag expression is correlated with a heterozygous genome composed of 16/13 repeats while low-tag expression is correlated to a homozygous genome (16/16 repeats).
This result was intriguing since minisatellites have been extensively correlated with human diseases (see Introduction) and it is known that the number of repeats could influence the level of expression of the nearby genes.
I thus hypothesized that the minisatellite could influence gene expression acting as an enhancer and therefore I performed a luciferase-based enhancer assay *in vitro* on the alleles. However, I didn't detect relevant activities neither a significant difference between the two alleles.
I then set up a genome-wide association study to look for other genomic loci associated to 13-repeats minisatellite conformation. Unfortunately, the analysis did not evidence any linkage.

Given that heterozygosity is more frequent in PD compared to HC according to the nanoCAGE cohort, I thought that minisatellite allele 13 could be a susceptibility marker for PD.
By genotyping more than one thousand PD individuals and HC we discovered that the frequency of heterozygosity is about 20% in both patients and controls. These results reject the hypothesis of allele 13 being linked to the pathology. Nonetheless, I observed that the allele 13 presented a much lower frequency in nanoCAGE-

analyzed HCs than in samples from DNA banks. Among the 60 HC from Associazione Donatori Sangue Trieste, only 7% were heterozygous. One potential explanation is that the frequency depends on the geographical area and ethnic composition of the sampled population.

The nanoCAGE study showed that the single HC individual with a 16/13 genotype presented the tag expression significantly lower than 16/13 PDs. While one sample is clearly insufficient for any scientific conclusion, I hypothesized that the presence of the allele 13 might influence gene expression in a disease-dependent manner. This would have tremendous consequences since the allele 13 may represent a stratification biomarker with diagnostic value.

Peripheral biomarkers are very much needed in PD since they represent a useful tool to stratify patients in order to choose/achieve the best personalized therapy (Precision Medicine). For example Chen-Plotkin et al., 2011 showed that epidermal growth factor in plasma may be a biomarker to predict cognitive impairment in PD.

To develop a gene expression assay for the repeat-containing transcript I needed to identify its full-length mRNA. To this purpose, I took advantage of the 3'RACE technique to isolate the complete cDNA towards the 3'end and then of the 5'RACE to confirm its TSS location in the minisatellite sequence.

Full-length cloning shows that the minisatellite sequence represents an alternative 5'end of a newly discovered Nprl3 isoform that we called TagNprl3.

Transcription does not start from the canonical initiator called Inr but from a less common core promoter sequence called TCT motif (YC+1TYTYY), which has been found to be specific for ribosomal protein coding genes and others involved in protein synthesis (Parry et al., 2010).

This result suggests TagNprl3 might be involved in protein synthesis.

I then validated the TagNprl3 transcript and designed a specific assay for its detection by quantitative TaqMan RT-PCR. I thus analyzed its expression in four independent sets of individuals: PD and HC of both 16/13 and 16/16 individuals.

TagNprl3 expression was higher in 16/13 PDs and HCs individuals revealing an allele-dependent high transcription rather than a disease-dependent one. Remarkably, minisatellite conformation does not influence Nprl3 canonical isoform 1 expression.

To our knowledge, this is the first time that a TSS is located in a minisatellite region, which variant in turn regulates its expression. A robust literature is available for

minisatellites that can influence gene expression when located in the promoter region, but that are not part of the mature mRNA. Among them, the transcribed minisatellite of the DRD4 gene is located in the third intron but its length does not influence the rate of expression or the function of the translated protein (see Introduction). In this case, TagNprl3 TSS is located in a minisatellite which is partially transcribed (it is part of the 5'UTR) and which length regulates its expression.

In this context I have found intriguing the recent findings on ATRX protein, a chromatin-remodelling factor that localizes throughout the genome at G-rich and CpG rich sequences including tandem repeats and CpG islands (Law et al., 2010). The protein is extensively studied since its mutations are associated to X-linked mental retardation with alpha thalassemia. Interestingly, mutant ATRX causes a downregulation of alpha globin genes due to its binding to a VNTR located within the cluster. In individuals carrying the same ATRX mutation, the degree of downregulation is proportional to the VNTR length.

Therefore, ATRX alters gene expression via an interaction with G-rich repeats while it remains unclear how the protein recognizes repeated sequences. Law et al., 2010 proposed that ATRX might recognize G-quadruplex (G4) structures that are likely to form when G-rich sequences are single stranded, as in DNA replication or transcription. The G4 structures may interfere with these processes and alter local gene expression while ATRX could help resolving such structures.

An interesting hypothesis to test in the future involves the influence on gene expression of ATRX binding to the minisatellite and assuming different binding affinities to alleles 16 or 13.

The next question I wanted to ask was whether TagNprl3 transcript is translated into protein. Therefore, I cloned its full-length sequence and expressed it *in vitro*. By using an anti Nprl3 polyclonal antibody I was able to detect a 60 kDa protein that corresponds to a N-terminal truncated Nprl3 isoform.

Nprl3 gene was discovered in 1995 by Vyas and colleagues that described its sequence and its evolutionary conserved position near the alpha globin cluster.

Although it is a widely expressed gene and it is highly conserved in metazoa, the function of the translated protein remained unclear for years, until 2009 when Neklesa & Davis demonstrated that Nprl3 forms a dimer with its homologous Nprl2

and together they take part to a complex responsible of mTORC1 inhibition. Nprl3 expression alterations have been correlated to mTOR pathway perturbation. Kowalczyk et al. in 2012 showed that Nprl3 ablation in mouse causes gestational death and the perturbation of protein synthesis and cell cycle genes, tightly regulated by mTOR. Recently, mTOR activation was found in patients with familial focal cortical dysplasia (FCD) as a consequence of Nprl3 gene mutations (Sim et al., 2015).

Nevertheless all the published works studied the effect of Nprl3 downregulation and nothing is known about the possible effects of its upregulation.

Since TagNprl3 expression is linked to genotype rather than to PD, I investigated the cellular source of this transcript, the presence of the translated protein and the biological consequences of TagNprl3 induction *in vitro*.

I thus demonstrated that TagNprl3 is expressed in RBC by performing blood fractionation followed by RNA and protein extraction. I detected TagNprl3 transcript by Real-Time TaqMan RT-PCR comparing high-expressing *versus* low-expressing samples and subsequently I detected its protein products *in vivo* by western blot, demonstrating that TagNprl3 is strongly upregulated in RBCs of 16/13 individuals.

By taking advantage of immunoprecipitation assays *in vitro*, we demonstrated that TagNprl3, like canonical Nprl3, forms a dimer with its partner Nprl2, despites lacking the first 78 aminoacids.

Interestingly, the longin domain (LD) at the N-terminal found by Levine et al. in 2013 is not necessary for this interaction since Nprl3 isoform 2, lacking this domain, maintain its ability to dimerize with Nprl2. The authors suggest that LDs could act as a platform for a GTPase involved in TORC1 signaling. If this is true, TagNprl3 could compete with canonical Nprl3.

By looking for differential function between TagNprl3 and canonical Nprl3, we first analysed mTORC1 downstream pathways with WB, discovering TaqNprl3 overexpression has no effects on autophagy neither on protein synthesis.

By performing BrdU assays, we demonstrated that TagNprl3 retains the ability of the full-length protein to inhibit proliferation, as in Lunardi et al. in 2009. They proposed that this is due to Nprl3 interaction with p73, a negative regulator of cell growth and involved in embryonic development, cell differentiation and tumor suppression.

Although TagNprl3 protein influence on mTORC1 activity must be proven, the fact that it is overexpressed in RBCs raises intriguing hypothesis on its function. Indeed

mTORC1 activity is crucial for RBCs, as demonstrated in mice: blood reticulocytes present high mTORC1 activity that could be modulated by iron availability and metabolism (Bayeva et al., 2012; Knight et al., 2014; La et al., 2013; Ohyashiki et al., 2009). Imbalances in mTORC1 function lead to different types of anemias with changes in RBC size, number and hemoglobin content (Knight et al., 2014); moreover, in humans the pharmacological activation of mTORC1 can ameliorate some types of hereditary anemias (Jaako et al., 2012; Payne et al., 2012).
Furthermore, a SNP at the Nprl3 gene has been found to be correlated to haemolysis in sickle cell anemia (Milton et al., 2012).
Nprl2 absence in mice severely compromises hematopoiesis (Dutchak et al., 2015) since it takes part to the GATOR1 protein complex that inhibits mTORC1, and it regulates the uptake and availability of cobalamin. This vitamin is required for methionine and *S*-Adenosyl methionin (SAM) synthesis that are part of important metabolic pathways for hematopoiesis (Klee, 2000; Koury and Ponka, 2004).
More generally, Nprl2 has been shown to be involved in DNA mismatch repair, cell cycle signalling, regulation of apoptosis and, importantly, is a candidate tumor suppressor gene. Nprl2 expression can be also a marker for malignancy (Otani et al., 2009).
On the other hand, the current understanding of Nprl3 function is essentially related to its participation to the protein complex GATOR1 inhibiting mTORC1 and to the studies investigating the effects of its absence or functional mutations.

In summary, future efforts should be focused on the understanding of the role of Nprl3 in cell homeostasis and the effects of its induction in RBCs. This work provides hints for Nprl3 protein function in blood and may suggest for the first time a link between mTOR and genomic polymorphisms in modifier genes. It will be interesting to study the association of these variants with hematopoiesis alterations and with cancers.
Even if we could not correlate minisatellite genotype and TagNprl3 expression with PD, it is possible that this marked induction in the blood of 16/13 individuals may influence the course of the pathology. In this context, it is particularly intriguing that the TagNprl3 binding partner Nprl2 is paying a fundamental role in Vitamin B12 and methionine metabolism, a pathway targeted by levodopa, the current symptomatic treatment of PD. It will be interesting to evaluate whether 16/16 and 16/13 possess

different responses to levodopa pharmacological treatment.

These results are included in a manuscript in preparation entitled “Discovery of a human VNTR allelic variant in Nprl3 gene intron that enhances its transcription in peripheral blood” (Maria Bertuzzi, Dave Tang, Raffaella Calligaris, Christina Vlachouli, Sara Finaurini, Stefano Goldwurm, Gilberto Pizzolato, Piero Carninci, Stefano Gustincich).

Bibliography

Agranat, L., Raitskin, O., Sperling, J., and Sperling, R. (2008). The editing enzyme ADAR1 and the mRNA surveillance protein hUpf1 interact in the cell nucleus. Proc. Natl. Acad. Sci. U. S. A. *105*, 5028–5033.

Alam, M., and Schmidt, W.J. (2002). Rotenone destroys dopaminergic neurons and induces parkinsonian symptoms in rats. Behav. Brain Res. *136*, 317–324.

Bar-Peled, L., Chantranupong, L., Cherniack, A.D., Chen, W.W., Ottina, K.A., Grabiner, B.C., Spear, E.D., Carter, S.L., Meyerson, M., and Sabatini, D.M. (2013). A Tumor suppressor complex with GAP activity for the Rag GTPases that signal amino acid sufficiency to mTORC1. Science *340*, 1100–1106.

Barroso, N., Campos, Y., Huertas, R., Esteban, J., Molina, J.A., Alonso, A., Gutierrez-Rivas, E., and Arenas, J. (1993). Respiratory chain enzyme activities in lymphocytes from untreated patients with Parkinson disease. Clin. Chem. *39*, 667–669.

Bayeva, M., Khechaduri, A., Puig, S., Chang, H.-C., Patial, S., Blackshear, P.J., and Ardehali, H. (2012). mTOR regulates cellular iron homeostasis through tristetraprolin. Cell Metab. *16*, 645–657.

Bennett, S.T., Wilson, A.J., Esposito, L., Bouzekri, N., Undlien, D.E., Cucca, F., Nisticò, L., Buzzetti, R., Bosi, E., Pociot, F., et al. (1997). Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. The IMDIAB Group. Nat. Genet. *17*, 350–352.

Berg, D., Godau, J., Seppi, K., Behnke, S., Liepelt-Scarfone, I., Lerche, S., Stockner, H., Gaenslen, A., Mahlknecht, P., Huber, H., et al. (2013). The PRIPS study: screening battery for subjects at risk for Parkinson's disease. Eur. J. Neurol. Off. J. Eur. Fed. Neurol. Soc. *20*, 102–108.

Bernet, A., Sabatier, S., Picketts, D.J., Ouazana, R., Morlé, F., Higgs, D.R., and Godet, J. (1995). Targeted inactivation of the major positive regulatory element (HS-40) of the human alpha-globin gene locus. Blood *86*, 1202–1211.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669–681.

Bessler, H., Djaldetti, R., Salman, H., Bergman, M., and Djaldetti, M. (1999). IL-1 beta, IL-2, IL-6 and TNF-alpha production by peripheral blood mononuclear cells from patients with Parkinson's disease. Biomed. Pharmacother. Bioméd. Pharmacothérapie *53*, 141–145.

Betarbet, R., Sherer, T.B., MacKenzie, G., Garcia-Osuna, M., Panov, A.V., and Greenamyre, J.T. (2000). Chronic systemic pesticide exposure reproduces features of Parkinson's disease. Nat. Neurosci. *3*, 1301–1306.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396–398.

Bjedov, I., Toivonen, J.M., Kerr, F., Slack, C., Jacobson, J., Foley, A., and Partridge, L. (2010). Mechanisms of life span extension by rapamycin in the fruit fly Drosophila melanogaster. Cell Metab. *11*, 35–46.

Blandini, F., Nappi, G., and Greenamyre, J.T. (1998). Quantitative study of mitochondrial complex I in platelets of parkinsonian patients. Mov. Disord. Off. J. Mov. Disord. Soc. *13*, 11–15.

Bongioanni, P., Mondino, C., Boccardi, B., Borgna, M., and Castagna, M. (1996). Monoamine oxidase molecular activity in platelets of parkinsonian and demented patients. Neurodegener. J. Neurodegener. Disord. Neuroprotection Neuroregeneration *5*, 351–357.

Bové, J., Martínez-Vicente, M., and Vila, M. (2011). Fighting neurodegeneration with rapamycin: mechanistic insights. Nat. Rev. Neurosci. *12*, 437–452.

Braak, H., Ghebremedhin, E., Rüb, U., Bratzke, H., and Del Tredici, K. (2004). Stages in the development of Parkinson's disease-related pathology. Cell Tissue Res. *318*, 121–134.

Brem, R.B., Yvert, G., Clinton, R., and Kruglyak, L. (2002). Genetic Dissection of Transcriptional Regulation in Budding Yeast. Science *296*, 752–755.

Brookes, K.J. (2013). The VNTR in complex disorders: the forgotten polymorphisms? A functional way forward? Genomics *101*, 273–281.

Brown, E.J., Albers, M.W., Shin, T.B., Ichikawa, K., Keith, C.T., Lane, W.S., and Schreiber, S.L. (1994). A mammalian protein targeted by G1-arresting rapamycin-receptor complex. Nature *369*, 756–758.

Cafferkey, R., Young, P.R., McLaughlin, M.M., Bergsma, D.J., Koltin, Y., Sathe, G.M., Faucette, L., Eng, W.K., Johnson, R.K., and Livi, G.P. (1993). Dominant missense mutations in a novel yeast protein related to mammalian phosphatidylinositol 3-kinase and VPS34 abrogate rapamycin cytotoxicity. Mol. Cell. Biol. *13*, 6012–6023.

Cai, C.Q., Zhang, T., Breslin, M.B., Giraud, M., and Lan, M.S. (2011). Both polymorphic variable number of tandem repeats and autoimmune regulator modulate differential expression of insulin in human thymic epithelial cells. Diabetes *60*, 336–344.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A.M., Taylor, M.S., Engström, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. *38*, 626–635.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Chen, Y.-Y., Lin, S.-Y., Yeh, Y.-Y., Hsiao, H.-H., Wu, C.-Y., Chen, S.-T., and Wang, A.H.-J. (2005). A modified protein precipitation procedure for efficient removal of albumin from serum. Electrophoresis *26*, 2117–2127.

Chen-Plotkin, A.S., Hu, W.T., Siderowf, A., Weintraub, D., Gross, R.G., Hurtig, H.I., Xie, S.X., Arnold, S.E., Grossman, M., Clark, C.M., et al. (2011). Plasma EGF levels predict cognitive decline in Parkinson's Disease. Ann. Neurol. *69*, 655–663.

Chiaruttini, C., Sonego, M., Baj, G., Simonato, M., and Tongiorgi, E. (2008). BDNF mRNA splice variants display activity-dependent targeting to distinct hippocampal laminae. Mol. Cell. Neurosci. *37*, 11–19.

Chio, C.L., Drong, R.F., Riley, D.T., Gill, G.S., Slightom, J.L., and Huff, R.M. (1994). D4 dopamine receptor-mediated signaling events determined in transfected Chinese hamster ovary cells. J. Biol. Chem. *269*, 11813–11819.

Clements, C.M., McNally, R.S., Conti, B.J., Mak, T.W., and Ting, J.P.-Y. (2006). DJ-1, a cancer- and Parkinson's disease-associated protein, stabilizes the antioxidant transcriptional master regulator Nrf2. Proc. Natl. Acad. Sci. U. S. A. *103*, 15091–15096.

Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., and Walters, L. (1998). New goals for the U.S. Human Genome Project: 1998-2003. Science *282*, 682–689.

Cookson, W., Liang, L., Abecasis, G., Moffatt, M., and Lathrop, M. (2009). Mapping complex disease traits with global gene expression. Nat. Rev. Genet. *10*, 184–194.

Courbon, F., Brefel-Courbon, C., Thalamas, C., Alibelli, M.-J., Berry, I., Montastruc, J.-L., Rascol, O., and Senard, J.-M. (2003). Cardiac MIBG scintigraphy is a sensitive tool for detecting cardiac sympathetic denervation in Parkinson's disease. Mov. Disord. Off. J. Mov. Disord. Soc. *18*, 890–897.

Dauer, W., and Przedborski, S. (2003). Parkinson's disease: mechanisms and models. Neuron *39*, 889–909.

Dehay, B., Bové, J., Rodríguez-Muela, N., Perier, C., Recasens, A., Boya, P., and Vila, M. (2010). Pathogenic lysosomal depletion in Parkinson's disease. J. Neurosci. Off. J. Soc. Neurosci. *30*, 12535–12544.

Denoeud, F., Vergnaud, G., and Benson, G. (2003). Predicting human minisatellite polymorphism. Genome Res. *13*, 856–867.

Dodson, M.W., and Guo, M. (2007). Pink1, Parkin, DJ-1 and mitochondrial dysfunction in Parkinson's disease. Curr. Opin. Neurobiol. *17*, 331–337.

Donadio, V., Incensi, A., Leta, V., Giannoccaro, M.P., Scaglione, C., Martinelli, P., Capellari, S., Avoni, P., Baruzzi, A., and Liguori, R. (2014). Skin nerve α-synuclein deposits: a biomarker for idiopathic Parkinson disease. Neurology *82*, 1362–1369.

Durinovic-Belló, I., Jelinek, E., Schlosser, M., Eiermann, T., Boehm, B.O., Karges, W., Marchand, L., and Polychronakos, C. (2005). Class III alleles at the insulin

VNTR polymorphism are associated with regulatory T-cell responses to proinsulin epitopes in HLA-DR4, DQ8 individuals. Diabetes *54 Suppl 2*, S18–S24.

Dutchak, P.A., Laxman, S., Estill, S.J., Wang, C., Wang, Y., Wang, Y., Bulut, G.B., Gao, J., Huang, L.J., and Tu, B.P. (2015). Regulation of Hematopoiesis and Methionine Homeostasis by mTORC1 Inhibitor NPRL2. Cell Rep. *12*, 371–379.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. Nature *452*, 423–428.

Flint, J., Tufarelli, C., Peden, J., Clark, K., Daniels, R.J., Hardison, R., Miller, W., Philipsen, S., Tan-Un, K.C., McMorrow, T., et al. (2001). Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. Hum. Mol. Genet. *10*, 371–382.

Foley, P., and Riederer, P. (2000). Influence of neurotoxins and oxidative stress on the onset and progression of Parkinson's disease. J. Neurol. *247 Suppl 2*, II82–II94.

Gasser, T. (2015). Usefulness of Genetic Testing in PD and PD Trials: A Balanced Review. J. Park. Dis. *5*, 209–215.

Gatt, J.M., Burton, K.L.O., Williams, L.M., and Schofield, P.R. (2015). Specific and common genes implicated across major mental disorders: a review of meta-analysis studies. J. Psychiatr. Res. *60*, 1–13.

Gebauer, F., and Hentze, M.W. (2004). Molecular mechanisms of translational control. Nat. Rev. Mol. Cell Biol. *5*, 827–835.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635–638.

Grada, A., and Weinbrecht, K. (2013). Next-Generation Sequencing: Methodology and Application. J. Invest. Dermatol. *133*, e11.

Gustincich, S., Sandelin, A., Plessy, C., Katayama, S., Simone, R., Lazarevic, D., Hayashizaki, Y., and Carninci, P. (2006). The complexity of the mammalian transcriptome. J. Physiol. *575*, 321–332.

Haas, R.H., Nasirian, F., Nakano, K., Ward, D., Pay, M., Hill, R., and Shults, C.W. (1995). Low platelet mitochondrial complex I and complex II/III activity in early untreated Parkinson's disease. Ann. Neurol. *37*, 714–722.

Halliwell, B. (1992). Reactive oxygen species and the central nervous system. J. Neurochem. *59*, 1609–1623.

Hansen, M., Chandra, A., Mitic, L.L., Onken, B., Driscoll, M., and Kenyon, C. (2008). A role for autophagy in the extension of lifespan by dietary restriction in C. elegans. PLoS Genet. *4*, e24.

Hartl, F.U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. Science *295*, 1852–1858.

Higgs, D.R. (2013). The Molecular Basis of α-Thalassemia. Cold Spring Harb. Perspect. Med. *3*.

Huang, R.S., Duan, S., Bleibel, W.K., Kistner, E.O., Zhang, W., Clark, T.A., Chen, T.X., Schweitzer, A.C., Blume, J.E., Cox, N.J., et al. (2007). A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. Proc. Natl. Acad. Sci. U. S. A. *104*, 9758–9763.

Hughes, A.J., Daniel, S.E., and Lees, A.J. (2001). Improved accuracy of clinical diagnosis of Lewy body Parkinson's disease. Neurology *57*, 1497–1499.

Hughes, J.R., Cheng, J.-F., Ventress, N., Prabhakar, S., Clark, K., Anguita, E., Gobbi, M.D., Jong, P. de, Rubin, E., and Higgs, D.R. (2005). Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. Proc. Natl. Acad. Sci. U. S. A. *102*, 9830–9835.

Infante, J., Prieto, C., Sierra, M., Sánchez-Juan, P., González-Aramburu, I., Sánchez-Quintana, C., Berciano, J., Combarros, O., and Sainz, J. (2015). Identification of candidate genes for Parkinson's disease through blood transcriptome analysis in LRRK2-G2019S carriers, idiopathic cases, and controls. Neurobiol. Aging *36*, 1105–1109.

Jaako, P., Debnath, S., Olsson, K., Bryder, D., Flygare, J., and Karlsson, S. (2012). Dietary L-leucine improves the anemia in a mouse model for Diamond-Blackfan anemia. Blood *120*, 2225–2228.

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S.T. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol. Cell. Biol. *14*, 116–127.

Johnson, S.C., Rabinovitch, P.S., and Kaeberlein, M. (2013). mTOR is a key modulator of ageing and age-related disease. Nature *493*, 338–345.

Kalia, L.V., and Lang, A.E. (2015). Parkinson's disease. Lancet Lond. Engl. *386*, 896–912.

Karlsson, M.K., Sharma, P., Aasly, J., Toft, M., Skogar, O., Sæbø, S., and Lönneborg, A. (2013). Found in transcription: accurate Parkinson's disease classification in peripheral blood. J. Park. Dis. *3*, 19–29.

Kaur, D., Yantiri, F., Rajagopalan, S., Kumar, J., Mo, J.Q., Boonplueang, R., Viswanath, V., Jacobs, R., Yang, L., Beal, M.F., et al. (2003). Genetic or pharmacological iron chelation prevents MPTP-induced neurotoxicity in vivo: a novel therapy for Parkinson's disease. Neuron *37*, 899–909.

Kedmi, M., Bar-Shira, A., Gurevich, T., Giladi, N., and Orr-Urtreger, A. (2011). Decreased expression of B cell related genes in leukocytes of women with Parkinson's disease. Mol. Neurodegener. *6*, 66.

Kim, S., Seo, J.-H., and Suh, Y.-H. (2004). Alpha-synuclein, Parkinson's disease, and Alzheimer's disease. Parkinsonism Relat. Disord. *10 Suppl 1*, S9–S13.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kindler, S., Wang, H., Richter, D., and Tiedge, H. (2005). RNA Transport and Local Control of Translation. Annu. Rev. Cell Dev. Biol. *21*, 223–245.

King, H.C., and Sinha, A.A. (2001). Gene expression profile analysis by DNA microarrays: promise and pitfalls. JAMA *286*, 2280–2288.

Klee, G.G. (2000). Cobalamin and folate evaluation: measurement of methylmalonic acid and homocysteine vs vitamin B(12) and folate. Clin. Chem. *46*, 1277–1283.

Knight, Z.A., Schmidt, S.F., Birsoy, K., Tan, K., and Friedman, J.M. (2014). A critical role for mTORC1 in erythropoiesis and anemia. eLife *3*, e01913.

Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., et al. (2006). CAGE: cap analysis of gene expression. Nat. Methods *3*, 211–222.

Koury, M.J., and Ponka, P. (2004). New insights into erythropoiesis: the roles of folate, vitamin B12, and iron. Annu. Rev. Nutr. *24*, 105–131.

Kowalczyk, M.S., Hughes, J.R., Babbs, C., Sanchez-Pulido, L., Szumska, D., Sharpe, J.A., Sloane-Stanley, J.A., Morriss-Kay, G.M., Smoot, L.B., Roberts, A.E., et al. (2012). Nprl3 is required for normal development of the cardiovascular system. Mamm. Genome Off. J. Int. Mamm. Genome Soc. *23*, 404–415.

Krige, D., Carroll, M.T., Cooper, J.M., Marsden, C.D., and Schapira, A.H. (1992). Platelet mitochondrial function in Parkinson's disease. The Royal Kings and Queens Parkinson Disease Research Group. Ann. Neurol. *32*, 782–788.

Kunz, J., Henriquez, R., Schneider, U., Deuter-Reinhard, M., Movva, N.R., and Hall, M.N. (1993). Target of rapamycin in yeast, TOR2, is an essential phosphatidylinositol kinase homolog required for G1 progression. Cell *73*, 585–596.

Kwon, S., Barbarese, E., and Carson, J.H. (1999). The cis-acting RNA trafficking signal from myelin basic protein mRNA and its cognate trans-acting ligand hnRNP A2 enhance cap-dependent translation. J. Cell Biol. *147*, 247–256.

La, P., Yang, G., and Dennery, P.A. (2013). Mammalian target of rapamycin complex 1 (mTORC1)-mediated phosphorylation stabilizes ISCU protein: implications for iron metabolism. J. Biol. Chem. *288*, 12901–12909.

Langston, J.W., Ballard, P., Tetrud, J.W., and Irwin, I. (1983). Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. Science *219*, 979–980.

Laplante, M., and Sabatini, D.M. (2009). mTOR signaling at a glance. J. Cell Sci. *122*, 3589–3594.

Laplante, M., and Sabatini, D.M. (2012). mTOR signaling in growth control and disease. Cell *149*, 274–293.

Larumbe Ilundáin, R., Ferrer Valls, J.V., Viñes Rueda, J.J., Guerrero, D., and Fraile, P. (2001). [Case-control study of markers of oxidative stress and metabolism of blood iron in Parkinson's disease]. Rev. Esp. Salud Pública *75*, 43–53.

Law, M.J., Lower, K.M., Voon, H.P.J., Hughes, J.R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M., Morris, A., et al. (2010). ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. Cell *143*, 367–378.

Lesch, K.P., Bengel, D., Heils, A., Sabol, S.Z., Greenberg, B.D., Petri, S., Benjamin, J., Müller, C.R., Hamer, D.H., and Murphy, D.L. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. Science *274*, 1527–1531.

Levine, T.P., Daniels, R.D., Wong, L.H., Gatta, A.T., Gerondopoulos, A., and Barr, F.A. (2013). Discovery of new Longin and Roadblock domains that form platforms for small GTPases in Ragulator and TRAPP-II. Small GTPases *4*, 62–69.

Li, Q. (2006). A Melanesian α-thalassemia mutation suggests a novel mechanism for regulating gene expression. Genome Biol. *7*, 238.

Li, Z., Okamoto, K.-I., Hayashi, Y., and Sheng, M. (2004). The Importance of Dendritic Mitochondria in the Morphogenesis and Plasticity of Spines and Synapses. Cell *119*, 873–887.

Lin, X., Cook, T.J., Zabetian, C.P., Leverenz, J.B., Peskind, E.R., Hu, S.-C., Cain, K.C., Pan, C., Edgar, J.S., Goodlett, D.R., et al. (2012). DJ-1 isoforms in whole blood as potential biomarkers of Parkinson disease. Sci. Rep. *2*, 954.

Liot, G., Bossy, B., Lubitz, S., Kushnareva, Y., Sejbuk, N., and Bossy-Wetzel, E. (2009). Complex II inhibition by 3-NP causes mitochondrial fragmentation and neuronal cell death via an NMDA- and ROS-dependent pathway. Cell Death Differ. *16*, 899–909.

Lotrich, F.E., and Pollock, B.G. (2004). Meta-analysis of serotonin transporter polymorphisms and affective disorders. Psychiatr. Genet. *14*, 121–129.

Lower, K.M., Hughes, J.R., De Gobbi, M., Henderson, S., Viprakasit, V., Fisher, C., Goriely, A., Ayyub, H., Sloane-Stanley, J., Vernimmen, D., et al. (2009). Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. Proc. Natl. Acad. Sci. U. S. A. *106*, 21771–21776.

Lunardi, A., Chiacchiera, F., D'Este, E., Carotti, M., Dal Ferro, M., Di Minin, G., Del Sal, G., and Collavin, L. (2009). The evolutionary conserved gene C16orf35 encodes a nucleo-cytoplasmic protein that interacts with p73. Biochem. Biophys. Res. Commun. *388*, 428–433.

Malagelada, C., Jin, Z.H., Jackson-Lewis, V., Przedborski, S., and Greene, L.A. (2010). Rapamycin protects against neuron death in in vitro and in vivo models of Parkinson's disease. J. Neurosci. Off. J. Soc. Neurosci. *30*, 1166–1175.

Mann, V.M., Cooper, J.M., Daniel, S.E., Srai, K., Jenner, P., Marsden, C.D., and Schapira, A.H. (1994). Complex I, iron, and ferritin in Parkinson's disease substantia nigra. Ann. Neurol. *36*, 876–881.

McCarroll, S.A. (2008). Extending genome-wide association studies to copy-number variation. Hum. Mol. Genet. *17*, R135–R142.

Migliore, L., Petrozzi, L., Lucetti, C., Gambaccini, G., Bernardini, S., Scarpato, R., Trippi, F., Barale, R., Frenzilli, G., Rodilla, V., et al. (2002). Oxidative damage and cytogenetic analysis in leukocytes of Parkinson's disease patients. Neurology *58*, 1809–1815.

Milton, J.N., Rooks, H., Drasar, E., McCabe, E.L., Baldwin, C.T., Melista, E., Gordeuk, V.R., Nouraie, M., Kato, G.R., Kato, G.J., et al. (2013). Genetic determinants of haemolysis in sickle cell anaemia. Br. J. Haematol. *161*, 270–278.

Molochnikov, L., Rabey, J.M., Dobronevsky, E., Bonucelli, U., Ceravolo, R., Frosini, D., Grünblatt, E., Riederer, P., Jacob, C., Aharon-Peretz, J., et al. (2012). A molecular signature in blood identifies early Parkinson's disease. Mol. Neurodegener. *7*, 26.

Mutez, E., Larvor, L., Leprêtre, F., Mouroux, V., Hamalek, D., Kerckaert, J.-P., Pérez-Tur, J., Waucquier, N., Vanbesien-Mailliot, C., Duflot, A., et al. (2011). Transcriptional profile of Parkinson blood mononuclear cells with LRRK2 mutation. Neurobiol. Aging *32*, 1839–1848.

Mutez, E., Nkiliza, A., Belarbi, K., de Broucker, A., Vanbesien-Mailliot, C., Bleuse, S., Duflot, A., Comptdaer, T., Semaille, P., Blervaque, R., et al. (2014). Involvement of the immune system, endocytosis and EIF2 signaling in both genetically determined and sporadic forms of Parkinson's disease. Neurobiol. Dis. *63*, 165–170.

Neklesa, T.K., and Davis, R.W. (2009). A genome-wide screen for regulators of TORC1 in response to amino acid starvation reveals a conserved Npr2/3 complex. PLoS Genet. *5*, e1000515.

Nicklas, W.J., Youngster, S.K., Kindt, M.V., and Heikkila, R.E. (1987). MPTP, MPP+ and mitochondrial function. Life Sci. *40*, 721–729.

Ohyashiki, J.H., Kobayashi, C., Hamamura, R., Okabe, S., Tauchi, T., and Ohyashiki, K. (2009). The oral iron chelator deferasirox represses signaling through the mTOR in myeloid leukemia cells by enhancing expression of REDD1. Cancer Sci. *100*, 970–977.

Orphanides, G., and Reinberg, D. (2002). A unified theory of gene expression. Cell *108*, 439–451.

Otani, S., Takeda, S., Yamada, S., Sakakima, Y., Sugimoto, H., Nomoto, S., Kasuya, H., Kanazumi, N., Nagasaka, T., and Nakao, A. (2009). The tumor suppressor NPRL2 in hepatocellular carcinoma plays an important role in progression and can be served as an independent prognostic factor. J. Surg. Oncol. *100*, 358–363.

Padeken, J., Zeller, P., and Gasser, S.M. (2015). Repeat DNA in genome organization and stability. Curr. Opin. Genet. Dev. *31*, 12–19.

Papapetropoulos, S., Adi, N., Ellul, J., Argyriou, A.A., and Chroni, E. (2007). A prospective study of familial versus sporadic Parkinson's disease. Neurodegener. Dis. *4*, 424–427.

Parkinson, J. (1817). An essay on the shaking palsy. 1817. J. Neuropsychiatry Clin. Neurosci. *14*, 223–236; discussion 222.

Parry, T.J., Theisen, J.W.M., Hsu, J.-Y., Wang, Y.-L., Corcoran, D.L., Eustice, M., Ohler, U., and Kadonaga, J.T. (2010). The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. Genes Dev. *24*, 2013–2018.

Pattabiraman, P.P., Tropea, D., Chiaruttini, C., Tongiorgi, E., Cattaneo, A., and Domenici, L. (2005). Neuronal activity regulates the developmental expression and subcellular localization of cortical BDNF mRNA isoforms in vivo. Mol. Cell. Neurosci. *28*, 556–570.

Payne, E.M., Virgilio, M., Narla, A., Sun, H., Levine, M., Paw, B.H., Berliner, N., Look, A.T., Ebert, B.L., and Khanna-Gupta, A. (2012). L-Leucine improves the anemia and developmental defects associated with Diamond-Blackfan anemia and del(5q) MDS by activating the mTOR pathway. Blood *120*, 2214–2224.

Pearson, T.A., and Manolio, T.A. (2008). How to interpret a genome-wide association study. JAMA *299*, 1335–1344.

Penn, A.M., Roberts, T., Hodder, J., Allen, P.S., Zhu, G., and Martin, W.R. (1995). Generalized mitochondrial dysfunction in Parkinson's disease detected by magnetic resonance spectroscopy of muscle. Neurology *45*, 2097–2099.

Petrozzi, L., Lucetti, C., Scarpato, R., Gambaccini, G., Trippi, F., Bernardini, S., Del Dotto, P., Migliore, L., and Bonuccelli, U. (2002). Cytogenetic alterations in lymphocytes of Alzheimer's disease and Parkinson's disease patients. Neurol. Sci. Off. J. Ital. Neurol. Soc. Ital. Soc. Clin. Neurophysiol. *23 Suppl 2*, S97–S98.

Pezawas, L., Meyer-Lindenberg, A., Drabant, E.M., Verchinski, B.A., Munoz, K.E., Kolachana, B.S., Egan, M.F., Mattay, V.S., Hariri, A.R., and Weinberger, D.R. (2005). 5-HTTLPR polymorphism impacts human cingulate-amygdala interactions: a genetic susceptibility mechanism for depression. Nat. Neurosci. *8*, 828–834.

Plessy, C., Bertin, N., Takahashi, H., Simone, R., Salimullah, M., Lassmann, T., Vitezic, M., Severin, J., Olivarius, S., Lazarevic, D., et al. (2010). Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. Nat. Methods *7*, 528–534.

Potashkin, J.A., Santiago, J.A., Ravina, B.M., Watts, A., and Leontovich, A.A. (2012). Biosignatures for Parkinson's disease and atypical parkinsonian disorders patients. PloS One *7*, e43595.

Ravikumar, B., Vacher, C., Berger, Z., Davies, J.E., Luo, S., Oroz, L.G., Scaravilli, F., Easton, D.F., Duden, R., O'Kane, C.J., et al. (2004). Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. Nat. Genet. *36*, 585–595.

Rockman, M.V., and Kruglyak, L. (2006). Genetics of global gene expression. Nat. Rev. Genet. *7*, 862–872.

Sabatini, D.M., Erdjument-Bromage, H., Lui, M., Tempst, P., and Snyder, S.H. (1994). RAFT1: a mammalian protein that binds to FKBP12 in a rapamycin-dependent fashion and is homologous to yeast TORs. Cell *78*, 35–43.

Sabers, C.J., Martin, M.M., Brunn, G.J., Williams, J.M., Dumont, F.J., Wiederrecht, G., and Abraham, R.T. (1995). Isolation of a Protein Target of the FKBP12-Rapamycin Complex in Mammalian Cells. J. Biol. Chem. *270*, 815–822.

Sabol, S.Z., Hu, S., and Hamer, D. (1998). A functional polymorphism in the monoamine oxidase A gene promoter. Hum. Genet. *103*, 273–279.

Salman, H., Bergman, M., Djaldetti, R., Bessler, H., and Djaldetti, M. (1999). Decreased phagocytic function in patients with Parkinson's disease. Biomed. Pharmacother. Biomed. Pharmacothérapie *53*, 146–148.

Santiago, J.A., and Potashkin, J.A. (2015). Network-based metaanalysis identifies HNF4A and PTBP1 as longitudinally dynamic biomarkers for Parkinson's disease. Proc. Natl. Acad. Sci. U. S. A. *112*, 2257–2262.

Santini, E., Alcacer, C., Cacciatore, S., Heiman, M., Hervé, D., Greengard, P., Girault, J.-A., Valjent, E., and Fisone, G. (2009). l-DOPA activates ERK signaling and phosphorylates histone H3 in the striatonigral medium spiny neurons of hemiparkinsonian mice. J. Neurochem. *108*, 621–633.

Schadt, E.E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P.Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C., et al. (2008). Mapping the genetic architecture of gene expression in human liver. PLoS Biol. *6*, e107.

Schapira, A.H., Cooper, J.M., Dexter, D., Jenner, P., Clark, J.B., and Marsden, C.D. (1989). Mitochondrial complex I deficiency in Parkinson's disease. Lancet Lond. Engl. *1*, 1269.

Scherzer, C.R., Eklund, A.C., Morse, L.J., Liao, Z., Locascio, J.J., Fefer, D., Schwarzschild, M.A., Schlossmacher, M.G., Hauser, M.A., Vance, J.M., et al. (2007). Molecular markers of early Parkinson's disease based on gene expression in blood. Proc. Natl. Acad. Sci. U. S. A. *104*, 955–960.

Shaikh, S., Ball, D., Craddock, N., Castle, D., Hunt, N., Mant, R., Owen, M., Collier, D., and Gill, M. (1993). The dopamine D3 receptor gene: no association with bipolar affective disorder. J. Med. Genet. *30*, 308–309.

Shaw, R.J. (2013). GATORs Take a Bite Out of mTOR. Science *340*, 1056–1057.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc. Natl. Acad. Sci. U. S. A. *100*, 15776–15781.

Shults, C.W., and Haas, R. (2005). Clinical trials of coenzyme Q10 in neurological disorders. BioFactors Oxf. Engl. *25*, 117–126.

Sim, J.C., Scerri, T., Fanjul-Fernández, M., Riseley, J.R., Gillies, G., Pope, K., van Roozendaal, H., Heng, J.I., Mandelstam, S.A., McGillivray, G., et al. (2015). Familial cortical dysplasia caused by mutation in the mTOR regulator NPRL3. Ann. Neurol.

Soreq, L., Israel, Z., Bergman, H., and Soreq, H. (2008). Advanced microarray analysis highlights modified neuro-immune signaling in nucleated blood cells from Parkinson's disease patients. J. Neuroimmunol. *201-202*, 227–236.

Spillantini, M.G., and Goedert, M. (2013). Tau pathology and neurodegeneration. Lancet Neurol. *12*, 609–622.

Strachan, T., and Read, A.P. (2004). Human Molecular Genetics 3 (Garland Science).

Stuurman, N., Meijne, A.M., van der Pol, A.J., de Jong, L., van Driel, R., and van Renswoude, J. (1990). The nuclear matrix from cells of different origin. Evidence for a common set of matrix proteins. J. Biol. Chem. *265*, 5460–5465.

Suzuki, H., FANTOM Consortium, Forrest, A.R.R., van Nimwegen, E., Daub, C.O., Balwierz, P.J., Irvine, K.M., Lassmann, T., Ravasi, T., Hasegawa, Y., et al. (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. Nat. Genet. *41*, 553–562.

Tabor, H.K., Risch, N.J., and Myers, R.M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat. Rev. Genet. *3*, 391–397.

Taira, T., Saito, Y., Niki, T., Iguchi-Ariga, S.M.M., Takahashi, K., and Ariga, H. (2004). DJ-1 has a role in antioxidative stress to prevent cell death. EMBO Rep. *5*, 213–218.

Tang, Z., Bereczki, E., Zhang, H., Wang, S., Li, C., Ji, X., Branca, R.M., Lehtiö, J., Guan, Z., Filipcik, P., et al. (2013). Mammalian target of rapamycin (mTor) mediates tau protein dyshomeostasis: implication for Alzheimer disease. J. Biol. Chem. *288*, 15556–15570.

Tanner, C.M. (1989). The role of environmental toxins in the etiology of Parkinson's disease. Trends Neurosci. *12*, 49–54.

Taylor, D.J., Krige, D., Barnes, P.R., Kemp, G.J., Carroll, M.T., Mann, V.M., Cooper, J.M., Marsden, C.D., and Schapira, A.H. (1994). A 31P magnetic resonance spectroscopy study of mitochondrial function in skeletal muscle of patients with Parkinson's disease. J. Neurol. Sci. *125*, 77–81.

Thenganatt, M.A., and Jankovic, J. (2014). Parkinson disease subtypes. JAMA Neurol. *71*, 499–504.

Tompkins, M.M., Basgall, E.J., Zamrini, E., and Hill, W.D. (1997). Apoptotic-like changes in Lewy-body-associated disorders and normal aging in substantia nigral neurons. Am. J. Pathol. *150*, 119–131.

Tongiorgi, E., Armellin, M., Giulianini, P.G., Bregola, G., Zucchini, S., Paradiso, B., Steward, O., Cattaneo, A., and Simonato, M. (2004). Brain-derived neurotrophic factor mRNA and protein are targeted to discrete dendritic laminas by events that trigger epileptogenesis. J. Neurosci. Off. J. Soc. Neurosci. *24*, 6842–6852.

Tufarelli, C., Hardison, R., Miller, W., Hughes, J., Clark, K., Ventress, N., Frischauf, A.M., and Higgs, D.R. (2004). Comparative Analysis of the α-Like Globin Clusters in Mouse, Rat, and Human Chromosomes Indicates a Mechanism Underlying Breaks in Conserved Synteny. Genome Res. *14*, 623–630.

Unneberg, P., and Claverie, J.-M. (2007). Tentative Mapping of Transcription-Induced Interchromosomal Interaction using Chimeric EST and mRNA Data. PLoS ONE *2*.

Valen, E., Pascarella, G., Chalk, A., Maeda, N., Kojima, M., Kawazu, C., Murata, M., Nishiyori, H., Lazarevic, D., Motti, D., et al. (2009). Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. Genome Res. *19*, 255–265.

Van Craenenbroeck, K., Clark, S.D., Cox, M.J., Oak, J.N., Liu, F., and Van Tol, H.H.M. (2005). Folding efficiency is rate-limiting in dopamine D4 receptor biogenesis. J. Biol. Chem. *280*, 19350–19357.

Van Tol, H.H.M., Bunzow, J.R., Guan, H.-C., Sunahara, R.K., Seeman, P., Niznik, H.B., and Civelli, O. (1991). Cloning of the gene for a human dopamine D4 receptor with high affinity for the antipsychotic clozapine. Nature *350*, 610–614.

Vernimmen, D. (2014). Uncovering enhancer functions using the α-globin locus. PLoS Genet. *10*, e1004668.

Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G., and Higgs, D.R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. EMBO J. *26*, 2041–2051.

Vernimmen, D., Marques-Kranc, F., Sharpe, J.A., Sloane-Stanley, J.A., Wood, W.G., Wallace, H.A.C., Smith, A.J.H., and Higgs, D.R. (2009). Chromosome looping at the human alpha-globin locus is mediated via the major upstream regulatory element (HS -40). Blood *114*, 4253–4260.

Vézina, C., Kudelski, A., and Sehgal, S.N. (1975). Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle. J. Antibiot. (Tokyo) *28*, 721–726.

Visanji, N.P., Marras, C., Hazrati, L.-N., Liu, L.W.C., and Lang, A.E. (2014). Alimentary, my dear Watson? The challenges of enteric α-synuclein as a Parkinson's disease biomarker. Mov. Disord. Off. J. Mov. Disord. Soc. *29*, 444–450.

Vyas, P., Vickers, M.A., Picketts, D.J., and Higgs, D.R. (1995). Conservation of position and sequence of a novel, widely expressed gene containing the major human alpha-globin regulatory element. Genomics *29*, 679–689.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390–394.

Westra, H.-J., and Franke, L. (2014). From genome to function by studying eQTLs. Biochim. Biophys. Acta *1842*, 1896–1902.

Yoshino, H., Nakagawa-Hattori, Y., Kondo, T., and Mizuno, Y. (1992). Mitochondrial complex I and II activities of lymphocytes and platelets in Parkinson's disease. J. Neural Transm. Park. Dis. Dement. Sect. *4*, 27–34.

Yu, J., and Russell, J.E. (2001). Structural and functional analysis of an mRNP complex that mediates the high stability of human beta-globin mRNA. Mol. Cell. Biol. *21*, 5879–5888.

Zhang, W., Phillips, K., Wielgus, A.R., Liu, J., Albertini, A., Zucca, F.A., Faust, R., Qian, S.Y., Miller, D.S., Chignell, C.F., et al. (2011). Neuromelanin activates microglia and induces degeneration of dopaminergic neurons: implications for progression of Parkinson's disease. Neurotox. Res. *19*, 63–72.

Zhang, Y., Yao, L., Liu, W., Li, W., Tian, C., Wang, Z.Y., and Liu, D. (2014). Bioinformatics analysis raises candidate genes in blood for early screening of Parkinson's disease. Biomed. Environ. Sci. BES *27*, 462–465.

Zhong, H., Beaulaurier, J., Lum, P.Y., Molony, C., Yang, X., Macneil, D.J., Weingarth, D.T., Zhang, B., Greenawalt, D., Dobrin, R., et al. (2010). Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. PLoS Genet. *6*, e1000932.