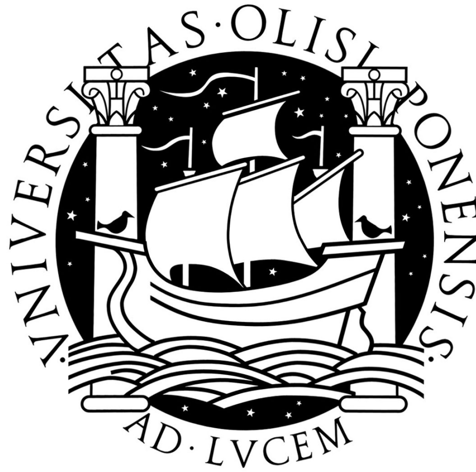
Universidade de Lisboa

Faculdade de Medicina de Lisboa



ZEBRAFISH:

A NEW MODEL OF PARKINSON'S DISEASE

Tomás Ribeiro da Silva Lopes da Fonseca

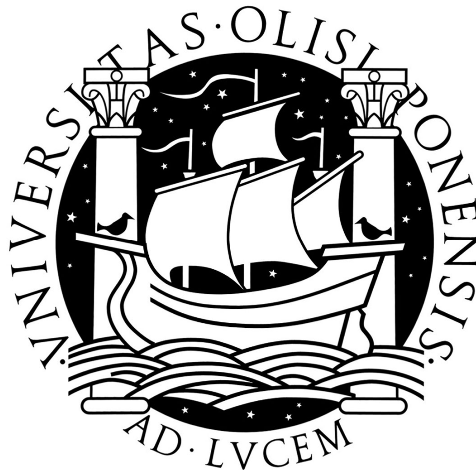
Mestrado em Neurociências

2010

A impressão desta dissertação foi aprovada pela Comissão Coordenadora do Conselho Científico da Faculdade de Medicina da Universidade de Lisboa, em reunião de 26 de Outubro de 2010 .

Universidade de Lisboa

Faculdade de Medicina de Lisboa



ZEBRAFISH:

A NEW MODEL OF PARKINSON'S DISEASE

Tomás Ribeiro da Silva Lopes da Fonseca

Mestrado em Neurociências, 2010

Dissertation supervised by

Prof. Doutor Tiago F. Outeiro, PhD

Doutora Ana Correia, PhD

Todas as afirmações efectuadas no presente documento são de exclusiva responsabilidade do seu autor, não cabendo qualquer responsabilidade à Faculdade de Medicina da Universidade de Lisboa pelos conteúdos nele apresentados.

Abstract

Parkinson´s disease (PD) is the second most common neurodegenerative disease after Alzheimer's disease, affecting around 2% of the population with more than 65 years.
The first gene implicated with PD was SNCA that encodes to alpha-synuclein. At the neuronal level, this protein was found in the Lewy Bodies, one of the pathologic hallmarks of the disease and it has been identified both in sporadic and familiar cases of PD. Its role in the disease is still unclear. Besides alpha-synuclein, other 15 genetic loci have been associated to PD.
To clarify the role of alpha-synuclein in PD, several *in vivo* models have been developed, expressing wild-type and PD-associated mutant forms of the protein. However, none of the existing models fully recapitulates all of the hallmarks of the disease. To circumvent this limitation, new models must be developed.
Zebrafish has been widely used as an animal model for studying several human disorders because, besides being a vertebrate, it has all the molecular background and bioimaging characteristics suitable for these studies.
Here, we started to generate a Zebrafish transgenic model for human alpha-synuclein.
In addition, through bioinformatic analysis, we found that four PD related genes encoding ATP13a2, Lrrk2, Synphilin-1, and Glucocerebrosidase, display high homology rates when compared to the human gene sequences. Moreover, despite a high homology in the entire sequence, specific conserved domains sequences are also highly homologous, suggesting a conserved functionality of the genes.
In situ hybridization of ATP13a2 revealed that, although initially expressed in whole embryo's body, it becomes more restricted to the brain area along the embryonic development.
The generation of this novel PD transgenic model will allow further studies to better understand the mechanisms underlying the disease pathology.
Taken together, these results will provide a novel and powerful system to explore more the molecular basis of PD and to open novel avenues for therapeutic intervention.

Keywords: Parkinson's disease, zebrafish, a-syn, ATP13a2, genetic model, bioinformatics

Resumo

A doença de Parkinson (PD) é a segunda doença neurodegenerativa mais comum depois da doença de Alzheimer, afectando cerca de 2% da população com mais de 65 anos.

SNCA foi o primeiro gene a ser associado a PD, que codifica para a alfa-sinucleína. A nível neuronal, esta proteína foi identificada nos corpos de Lewy, uma das principais características da doença, tanto em casos esporádicos como em casos familiares de Parkinson. O seu papel na doença ainda não foi desvendado. Além da alfa-sinucleína outros genes têm sido associados a casos familiares da PD.

Para esclarecer a sua função, têm sido desenvolvidos vários modelos *in vivo*, expressando as formas nativas e mutadas da proteína. No entanto nenhum foi capaz de recapitular todas as características da doença, especialmente em termos fenotípicos e comportamentais. Este facto levou à necessidade de gerar novos modelos.

Zebrafish tem sido amplamente utilizado como modelo animal e, para além de ser um vertebrado, apresenta as ferramentas moleculares e de bioimagiologia necessárias para ser utilizado como um modelo de doenças neurodegenerativas.

Neste trabalho, iniciámos a criação de um modelo de zebrafish transgénico, expressando alfa-sinucleína humana. Além disso, através de análise bioinformática, concluimos que quatro genes relacionados com PD: ATP13a2, LRRK2, sinfilina-1 e glucocerebrosidase apresentam uma elevada taxa de homologia em comparação com os respectivos genes humanos. Esta homologia estende-se aos domínios funcionais que são idênticos entre as sequências indicando uma funcionalidade conservada dos genes.

Através da técnica que hibridização *in situ*, demonstrámos que o gene ATP13a2 tem uma expressão ubíqua ao longo do corpo de Zebrafish durante as fases iniciais do seu desenvolvimento embrionário, sendo essa expressão mais localizada no cérebro em estadios mais avançados.

Em conjunto, estes resultados vão abrir uma nova perspectiva sobre a etiologia da PD.

Palavras-chave: Doença de Parkinson, Zebrafish, a-syn, ATP13a2, modelo genético, Bio-informática

List of abbreviations

a-syn	Alpha-synuclein
Blast	Basic local alignment search tool
bp	Base pairs
dpf	Days post fertilization
GBA	Glucocerebrosidase
GFP	Green fluorescent protein
hpf	Hours post fertilization
IRES	Internal ribosome entry site
kDa	kilo Dalton
LB	Lewy bodies
Lrrk2	Leucine-rich repeat kinase 2
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine
PD	Parkinson's disease
PINK-1	PTEN-induced putative kinase 1
PTEN	Phosphatase and tensin homolog
SMART	Simple modular architecture research tool
SNCAIP	Alpha-synuclein interacting protein
Tg	Transgenic
Wt	Wild-type

Table of Contents

1. Introduction…1
1.1 Parkinson's disease…1
1.2 PD associated genes…3
1.2.1 Alpha-synuclein…4
1.2.2 LRRK2…5
1.2.3 ATP13a2…6
1.2.4 Synphilin-1…7
1.2.5 Glucocerebrosidase…7
1.3 Zebrafish as a model organism in PD…8
2. Objectives…11
3. Methods…12
3.1 Construct Development…12
3.2 Microinjection…14
3.3 Zebrafish maintenance…15
3.4 Blasting…15
3.5 Phylogenetic tree and functional anaylisis…15
3.6 In-situ probe design…16
3.7 Whole-mount In situ Hybridization…17
3.8 Epifluorescence Microscopy…17
4. Results…18
4.1 Generating the transgenic zebrafish lines…18
4.2 Bioinformatic analysis revealed high degree of conservation of PD-associated genes between Human and Zebrafish…21
4.2.1 Zebrafish has one orthologue for Human ATP13a2-5…21
4.2.2 Human and Zebrafish Lrrk2 display 66% similarity…27
4.2.3 Functional Domain of SNCAIP is 93% similar between species…32
4.2.4 Glucocerebrosidase in zebrafish has the same functional domain found in humans…34
4.3 ATP13a2 is expressed during Zebrafish embryonic development…35

5. Discussion……………………………………………………………………………………………………37

5.1 Transgenic zebrafish as a model of PD…………………………………………………………37

5.2 PD genes in Zebrafish…………………………………………………………………………….38

6. Conclusion……………………………………………………………………………………………………41

7. Future Work………………………………………………………………………………………………….42

8. Acknowledgments………………………………………………………………………………………..43

9. References…………………………………………………………………………………………………..44

10. Annexes…………………………………………………………………………………………………….49

Annexe 1……………………………………………………………………………………………………49

Annexe 2……………………………………………………………………………………………………50

Index of Images

Figure 1. Spatial location of the *substantia nigra* in human brain…………………………………………….1

Figure 2. a-syn protein domains………………………………………………………………………………………4

Figure 3. Human Lrrk2 functional domains…………………………………………………………………………6

Figure 4. Constructs created for the generation of transgenic zebrafish………………………………12

Figure 5. Tol2 tranposase technique…………………………………………………………………………….13

Figure 6. PD transgenic Zebrafish protein fluorescence………………………………………………………20

Figure 7. F1 positive Generation…………………………………………………………………………………21

Figure 8. Multiple protein sequence alignment of human ATP13a2-5 and the orthologue zgc_136762 in Zebrafish. ………………………………………………………………………………………………..25

Figure 9. Phylogenetic tree including the human ATP13a -2 to -5 and the orthologue in Zebrafish. ………26

Figure 10. Side by side comparison of Human and Zebrafish ATP13a2…………………………………27

Figure 11. Multiple protein sequence alignment of Human and Zebrafish Lrrk2 orthologues…27

Figure12. Human and Zebrafish Lrrk2 proteins……………………………………………………………31

Figure 13. Alignment of the Human and Zebrafish Synphilin-1 sequences……………………………33

Figure 14. Human and Zebrafish synphilin-1………………………………………………………………33

Figure15. Multiple protein sequence alignment of Human GBA and its orthologue in Zebrafish………34

Figure 16. Human and Zebrafish GBA protein sequence homology and functional domain…….35

Figure 17. ATP13a2 expression during embryonic development…………………………… ……………36

Index of Tables

Table 1 - Parkinson Disease associated genes……………………………………………………………………………3

Table 2 - Identity and Similarity in PD genes between human and Zebrafish…………………………40

Table 3 - Identity and similarity of Functional Domains of studied genes………………………………40

1. Introduction

1.1 **Parkinson´s Disease**

Parkinson's disease (PD) was first described in 1817 by James Parkinson and is the second most common neurodegenerative disorder, after Alzheimer´s disease. PD is characterized by motor and non-motor features. The most common motor symptoms are tremor, rigidity, akinesia (reduction in movement) and bradykinesia (slowed movement) that occur mainly due to the degeneration of dopaminergic neurons, involved in the movement coordination and located in the *substantia nigra pars compacta* (Figure 1). However, in later stages of the disease it affects other brain regions [1, 2]. The degeneration of these wider circuits in the brain is responsible for the non-motor features of the disease, such as cognitive decline, depression and, in some cases, hallucination episodes.

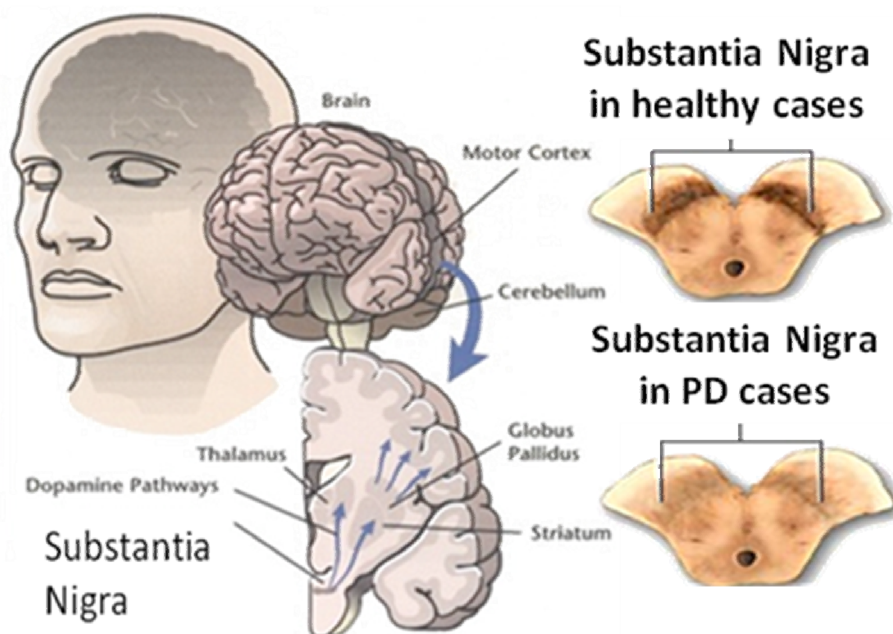


Figure 1. Spatial location of the *substantia nigra* in human brain. In PD patients it is visually possible to detect a depletion of dopaminergic neurons in the *substantia nigra* which is responsible for the majority of the motor symptoms found in the disease. Adapted from http://adam.about.com/reports/Parkinson-s-disease.htm and http://nihseniorhealth.gov/parkinsonsdisease/whatcausesparkinsonsdisease/02.html

The majority of the PD cases are sporadic and only 2% of the cases are familiar, being associated to specific gene mutations. Phenotypically, both PD forms are very similar regarding motor symptoms, suggesting that the insult responsible for the disease development and progression may be identical in both cases [3, 4].

The major pathologic hallmark of PD, besides the degeneration of dopaminergic neurons, is the presence of cytoplasmic protein inclusions named Lewy Bodies (LBs) that can be found in the remaining surviving neurons. These inclusions are mainly constituted by alpha-synuclein (a-syn) which was the first protein to be associated to the disease. The role of these bodies is still unknown, however it is believed that they may present a protective effect in the disease, by sequestering dysfunctional and toxic protein species responsible for neurodegeneration [5]. In agreement, there is an increasing hypothesis suggesting a toxic role at a cellular level in PD via oligomeric species formation of a-syn [6, 7]. Those intermediate species might be more toxic than the larger protein inclusions, whose formation represents a survival strategy to protect neurons from the toxicity induced by a-syn intermediate oligomers.

There are three known a-syn mutations linked to PD that are the result of missense alterations: A30P, E46K and A53T. The association of a-syn with PD was first identified due to the A53T mutation, which promotes autosomal dominant familial PD [8]. At a biochemical glance, this mutation is characterized by a disruption of the ubiquitin proteasome system and by a marked accumulation of autophagic-vesicular structure [9] .

The A30P mutation abolishes the ability of a-syn to bind to lipid vesicles [10] and in yeast this mutated protein is found in cytoplasm and not in the plasma membrane as the wt and A53T proteins [11]. More recently, it was found that this mutation stabilizes a-syn against proteasomal degradation triggered by heme oxygenase-1 overexpression [12].

Little is known about the third discovered mutation, E46K, at the exception that it promotes amyloid fibril formation [13] and leads to an atrophy of the substantia nigra and LBs formation [14].

Interestingly, triplication of the region that contains the a-syn gene and flanking regions was identified in PD and dementia with LBs [15].

Besides PD, a-syn is the common divisor in several neurodegenerative diseases called synucleinopathies, like Multiple System Atrophy and Dementia with LBs.

1.2 **PD-associated genes**

Since 1998, with the discovery of a-syn, several genes have been associated to PD: PARKIN, ubiquitin carboxyl-terminal esterase L1 (UCHL1), phosphatase and tensin homolog (PTEN)-induced putative kinase 1 (PINK1), DJ-1, Leucine-rich repeat kinase 2 (Lrrk2), ATP13a2, synphilin-1, glucocerebrosidase among others (Table 1).

Table 1. Known genetic loci linked to Parkinson's disease

mutation	gene	locus	Lewy Bodies	onset/age	inheritance
Park 1	a-syn	4q21	yes	40s	AD
Park 2	Parkin	6q25	no	20s	AR
Park 3	?	2p13	yes	60s	AD
Park 4	a-syn	4q21	yes	30s	AD
Park 5	UCH-L1	4p15	yes	50s	AD
Park 6	PINK1	1p35	?	30s	AR
Park 7	DJ1	1p36	?	30s	AR
Park 8	Lrrk2	12p	?	40s	AD
Park 9	ATP13A2	1p36	?	10s	AR
Park 10	?	1p32	?	50s	?
Park 11	GIGYF2	2q36-37	?	late	AD
Park 12	?	Xq21-q25	?	?	X-chromosome
Park 13	HTRA2	2p12	?	50s	?
Park 14	PLA2G6	18q11	?	?	AR
Park 15	FBXO7	22q12-q13	?	?	AR
Park 16	?	1q32	?	?	?

1.2.1 Alpha-synuclein

a-syn belongs to a family of three distinct genes, a-, β- and γ-synuclein which have only been found in vertebrates [16]. These proteins show a highly conserved alpha-helical-rich amino-terminal domain that is mainly unfolded in solution but can shift to an α-helical conformation upon interaction with lipid structures.

The carboxy-terminal domain is rich in proline and acidic residues like glutamate and aspartate. The central hydrophobic domain of a-syn contains a specific domain known as NAC (non-Aβ component of AD amyloid). This region has been implicated in Alzheimer's disease as the second main component of amyloid plaques. It is responsible for specific protein-protein interactions and is believed to contribute to an a-syn conformation change towards an aggregation-prone β-sheet structure [17] (Figure 2).

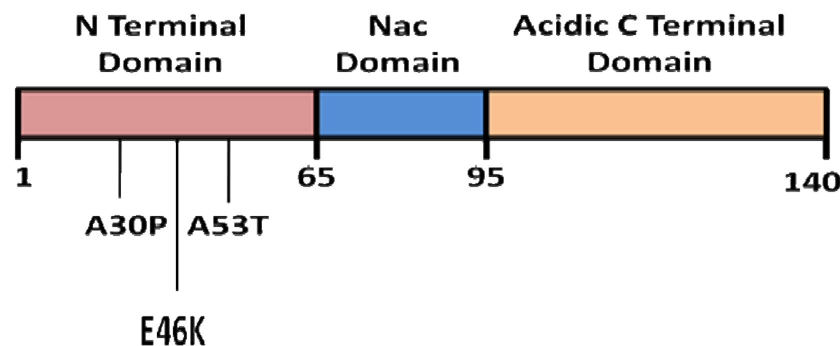


Figure 2. a-syn protein domains. The NAC domain is responsible for the protein structure instability, intrinsically prone to be unfolded. The three mutations are found in the N-terminal domain.

a-syn is a small 140 amino-acid protein with a molecular weight of 14.5 kDa, abundantly expressed in the brain, which is enriched in neural pre-synaptic terminals. Its role in normal or pathological conditions is still unclear. Although its function remains poorly unknown, a-syn has been associated to vesicular trafficking since its cytoplasm accumulation inhibits the vesicular traffic between the endoplasmic reticulum and the Golgi complex [18]. This protein has been also associated with the neuronal signaling in several steps: vesicle reserves, where the absence of the protein leads to a decrease in the reserve pool of synaptic vesicles [19, 20];

Vesicle release: in some cell lines, the a-syn overexpression impairs the release of neurotransmitters [21]. Moreover, it has been shown that a-syn plays a role in the lipid transport and in the biogenesis of the synaptic membrane. It is thought that its structural characteristics (amphipathic α-helical domains) are reminiscent to those of the class A2 apolipoproteins, responsible for lipid molecules trafficking [22]. Also in vesicle recycling since a-syn is implicated in the endocytic recycling of synaptic vesicles, following neuronal stimulation by interaction with polyunsaturated fatty acids[23].
In the disease, a-syn has also been implicated in cell death due to the toxicity exhibited by its oligomeric species [6, 24, 25].

1.2.2 Lrrk2

Lrrk2 is a 2527 amino-acid protein which was first associated with PD in 2004 and is now considered as the most important autosomal-dominant origin of PD [26]. Lrrk2 is a large protein (286kDa) with several predicted conserved domains (Figure 3): a Leucine-rich repeat (LRR), a WD40 domain, a mitogen-activated protein kinase kinase kinase domain (MAPKKK), a Ras in complex proteins (ROC), a C-terminal of Roc (COR) domain, and an Ankaryne repeat.
The LRR, WD40 and Ankaryne regions can be found in several other proteins and they are involved with different molecules interactions such as cytoskeleton proteins, transcription factors, and signaling and cell cycle regulators. COR and ROC are found together in ROCO family of GTPases. The function of the former domain is still unknown but the latter might regulate vesicle formation, actin- and tubulin-dependent vesicle movement, and membrane fusion, a role similar to the Ras-associated binding-GTPase (Rab) family of proteins [27]. The MAPKKK domain of Lrrk2 belongs to the super-family of serine/threonine and tyrosine kinases, which have an important role in cell survival and cell death pathways in response to intracellular or extracellular stress [28].
There are several mutations identified in Lrrk2, but only few were linked to a PD pathogenic effect: GS2019S, I2020T, both in the MAPK domain [29, 30], R1441C and

R1441G in the ROC domain, and Y1699C, that is found in the COR domain. The first one is responsible for 0,5-2,0% of the sporadic cases of PD .

Figure 3. Human Lrrk2 functional domains. Lrrk2 is the biggest protein linked to PD and is responsible for the majority of the familial cases of the disease. It has several functional domains: Ankyrin (ANK), Leucine-rich repeat (LRR), Ras in complex proteins (ROC), a C-terminal of Roc (COR), mitogen-activated protein kinase kinase kinase domain (MAPKKK), and a WD40 domain.

This protein can be found in brain, lungs, kidney, spleen and leucocytes [31-35]. In the rat brain it can be found at high concentrations (both mRNA and protein levels) in the dorsal striatum and cortex and also at lower levels in the *substantia nigra[36]*.

Inside the cell, Lrrk2 is present in the cytoplasm [37] and associated with the mitochondria [38, 39]. Moreover, it can also be found in lipid rafts on the cellular membrane of organelles like golgi, endoplasmic reticulum and synaptic vesicles. In fact, Lrrk2 was suggested to play a role in the regulation of the vesicle synthesis, transport and secretion and also in the regulation of membranous structures [40].

Lrrk2 has both a GTPase and a kinase domain, exhibiting a dual enzymatic activity. Lrrk2 GTPase activity seems to regulate its kinase activity. Mutations in the first domain promotes its kinase activity [39, 41] while the opposite does not have an effect on the GTPase activity[42].

1.2.3 ATP13a2

In 2006, *ATP13A2* was initially identified as the gene responsible for Kufor-Rakeb syndrome in two families. The disease is characterised by sub acute, juvenile-onset, levodopa-responsive parkinsonism, pyramidal signs, dementia and a supranuclear gaze palsy which consists in the inability to look in a particular direction.

ATP13A2 belongs the P_5 subfamily of P-type transport ATPases, a transporter family with unknown substrate specificity [43]. A more recent phylogenetic study

shows the relationship between the different orthologs of these proteins and the presence of duplications of *ATP13A2* in mice and humans, but not in lower vertebrates [44].

Two loss-of-function mutations in this protein were found in a early on-set hereditary parkinsonism case [45].

1.2.4 Synphilin-1

Synphilin-1, also known has a-syn interacting protein (SNCAIP), is also present in LBs [46]. The interaction with a-syn has been seen *in vivo* [47] an the co-expression of the two proteins in cell culture, were enough to form LB-like inclusions [48].

Besides a-syn, synphilin-1 also interacts with parkin[49] and Lrrk2 [50] suggesting that it has an important role both in sporadic and familial cases of PD.

The gene has also been found mutated (R621C) in two different PD cases [51].

1.2.5 Glucocerebrosidase

Mutations in Glucocerebosidase (GBA) are found in Gaucher's disease. These mutations reduce or eliminate the activity of beta-glucocerebrosidase in cells. As a result, glucocerebroside is not properly processed. This protein can accumulate in macrophages and damage the spleen, liver, bone marrow, and other organs. More recently, this protein has been linked to Parkinson [52]. Some of the characteristics found in Guacher's disease are identical to PD patients: tremor, bradykinesia, rigidity and often cognitive decline [52, 53]. In 2004, has been proven that PD patients have higher rates than expected of GBA in mutated form [54].

1.3 Zebrafish as a model organism in PD

To better understand the chemical, physiological and pathological effects of a-syn, *in vivo* experiments with animal models overexpressing the normal and mutant variants of this protein have been generated. Previous experiments using invertebrate models like *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster* proven to be useful in the study of PD [55].

Saccharomyces cerevisiae started to be used as a simple and versatile tool in high-throughput analysis, allowing researchers to understand the main mechanisms underlying a-syn-mediated cellular degeneration. The first genetic screen in yeast identified 86 yeast deletion mutants with enhanced a-syn toxicity. Several of these mutants lacked functions involved in lipid metabolism, vesicular transport, ubiquitin proteasome system, and defenses against oxidative stress or mitochondrial activities. [56].

Transgenic models in *Caenorhabditis elegans* and *Drosophila melanogaster* showed that a-syn overexpression led to the loss of dopaminergic neurons, resulting in motor coordination defects. In *D. melanogaster* this is an age-dependent effect [55, 57].

Nevertheless, there is a great disadvantage in invertebrate models of PD which is the different nervous system complexity compared to higher species. Moreover, these models lack endogenous a-syn [58].

On the other hand, several lines of transgenic mice have also been generated based on toxins and genetic approaches. Several toxins like 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), paraquat, epoxomicin, rotenone, 6-hidroxy dopamine have been proven to promote a parkinson-like behavior. MPTP, paraquat and rotenone are mitochondrial complex I inhibitors although they act through different mechanisms. The first two are specific for dopaminergic neurons while rotenone has a systemic effect. Epoxomicin is a proteossomal inhibitor being able to recapitulate most of the PD key hallmarks. In terms of genetic approaches, both overexpression or knock-out of specific PD related genes like a-syn, DJ-1, Parkin and Lrrk2 have been performed (reviewed at [59]).

The phenotypic results, however, do not resemble all the PD features. In some cases, it was evidenced a neuronal accumulation of a-syn [60] while in other models only few symptoms were attributed to the neurodegeneration of motor neurons [61].

To overcome this issue, *Danio rerio*, commonly designated as zebrafish, is now being used as an intermediate model between *D. melanogaster* and mice. This model has the advantage of being a vertebrate and therefore is phylogenetically closer to humans than drosophila, easier to maintain than mice, with a short life cycle and, importantly, presents a nervous system similar to humans[62].

In general, animal models must present certain features: a small life cycle, low maintenance cost, large number of descendants, and an easy manipulation. Zebrafish share all these aspects together with the advantage of being less expensive than mice. Moreover, due to its optical transparency during the different embryo stages, with the use of fluorescent reporters it allows *in vivo* gene expression and different cellular processes monitoring [63].

To develop a neurodegenerative model, the following properties in the regulatory elements must be verified: the gene/protein expression must occur in a controllable range of neurons, to investigate their role in different neurons sub-populations; the gene/protein expression cannot lead to an early death phenotype, since the main goal is to study their late effects in neurodegeneration; the animal must live until adulthood; the gene/protein expression must persist until adult age and finally the gene/protein expression must lead to a visible phenotype [64].

Zebrafish is thus an ideal organism to study neurodegeneration. Being a vertebrate, the basic organization and division of the nervous system is similar to other vertebrates, including humans [62]. The zebrafish central nervous system (CNS) contains specialized neuronal populations of direct relevance to human neurodegenerative diseases, for example dopaminergic neurons [65], cerebellar Purkinje cells [66] and motor neurons [67]. In addition, CNS also contains the human homologues of oligodendrocytes and astrocytes, which may play central roles in neurodegeneration through critical neuro-glial interactions. Therefore, Zebrafish can be used as a model system for neurodegenerative diseases.

Interestingly, zebrafish expresses several highly conserved and engodenous genes associated with human PD, including DJ-1 [68], Parkin [69], Pink1 [70], UCH-L1 [71] and three human synuclein orthologues also known as *sncga*, *sncgb*, and *sncb* [72].

The dopaminergic system in zebrafish homolog to the human one is located in the ventral diencephalon [73]. Treatment with MPTP showed a direct effect on these diencephalic dopaminergic neurons. This results in a loss of diencephalic dopaminergic neurons and an aberrant swimming pattern, indicating that MPTP may lead to zebrafish motor impairment, a known PD feature [74]. Therefore, Zebrafish should be a good model to study this disease and the generation of transgenic lines is the next step to achieve this goal.

Besides all the advantages of Zebrafish, it also has some disadvantages when compared to other models. They aren't mammals, so they are not as closely related to human as mice, therefore, all the new discoveries must later be verified in a mammal model. When compared to D. melanogaster, the number of genetic-based tools are much less, making fly a better model for those type of studies.

Here, we propose to develop transgenic lines overexpressing a-syn, by genome insertion in the zebrafish so it can be transmitted to the next generations.

2. Objectives

The main goal of this project was to create the first a-syn transgenic zebrafish line using human wild-type a-syn and GFP. We chose the Tyrosine Hydroxylase (TH) promoter to drive expression of the gene specifically in dopaminergic neurons.

A secondary objective was to investigate the presence of homologues genes of PD in zebrafish through bioinformatic analysis.

We believe that this model will recapitulate central aspects of PD, constituting a powerful *in vivo* model to study the molecular basis of the disease.

3. Methods

To generate the transgenic zebrafish lines:

3.1 **Development of the constructs**

The following constructs were generated: THrat:IRES:GFP, THrat:a-syn:IRES:GFP and THzeb:GFP (Figure 4) in pminitol2 vector (Annex1). The IRES:GFP sequence was obtain from the pIRES-hrGFP-1, a vector from stratagene (Annex 2).

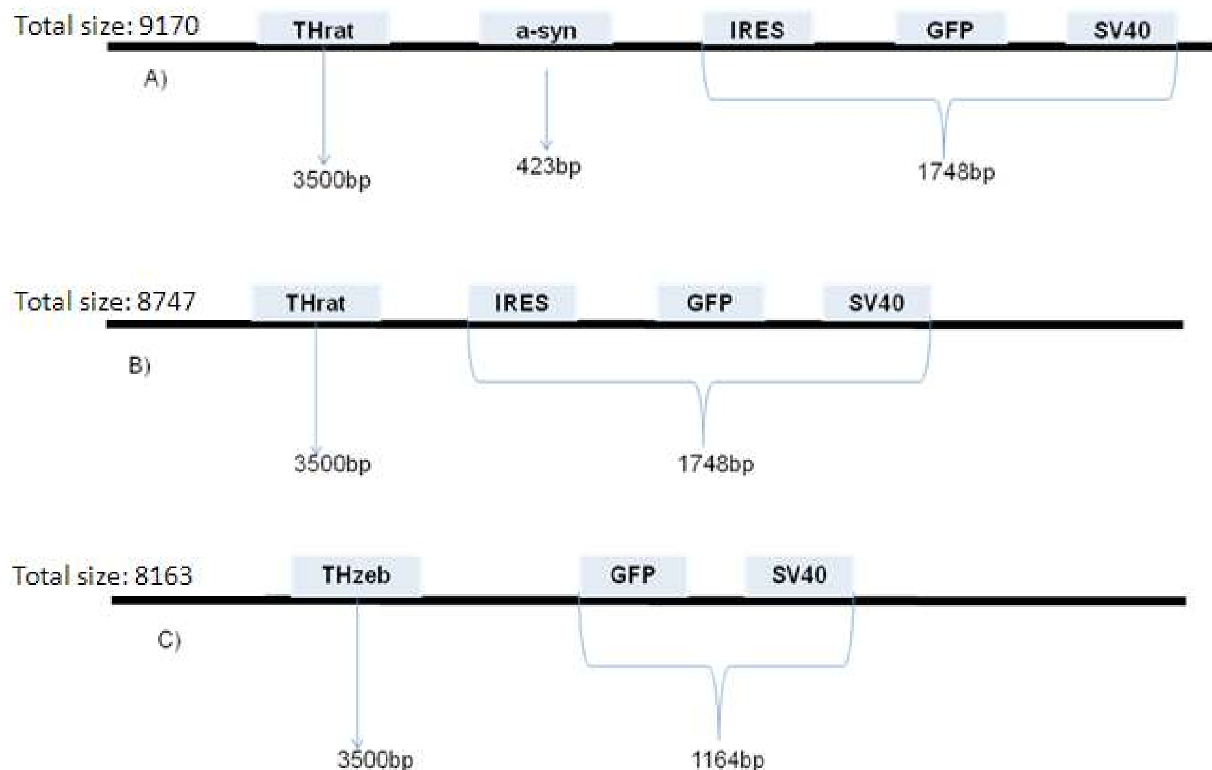


Figure 4. Constructs created for the generation of transgenic zebrafish. The constructs are on pMinitol vector. **A)** THrat:a-syn:IRES:GFP with the total size of 9170bp. **B)** THrat:IRES:GFP with 8747bp. **C)** The control construct Thzeb:GFP is the smaller with 8163bp.

The constructs were linearized and then co-injected with the mRNA transposase. There are two main techniques to achieve the insertion of a construction for the generation of transgenic lines: the use of the Tol2 transposase technique or the pseudotype retroviral technique.

The Tol2 transposase system is based on the use of a specific vector (pminitol2) which has two tol2 sequences at its extremities and a tol2 transposase mRNA. Tol2 element has 4.7 kb in length and contains a gene encoding a transposase protein [75]. The transcribed mRNA encodes for the transposase protein fully functional that catalyzes the transposition of a non-autonomous Tol2 construct [75, 76]. This transposition in the genome can be used up to 11 kb sequences with a single-copy insertion [77]. The rate of success ranges from 20% to 70% concerning the germline transmitting founder fish [77, 78]. The co-injection of the vector with the construct and the transposase mRNA will result in the translation of mRNA transposase which might catalyze the construct excision from the pminitol2 vector leading to a stable integration of the excised DNA in the genome (Figure 5).

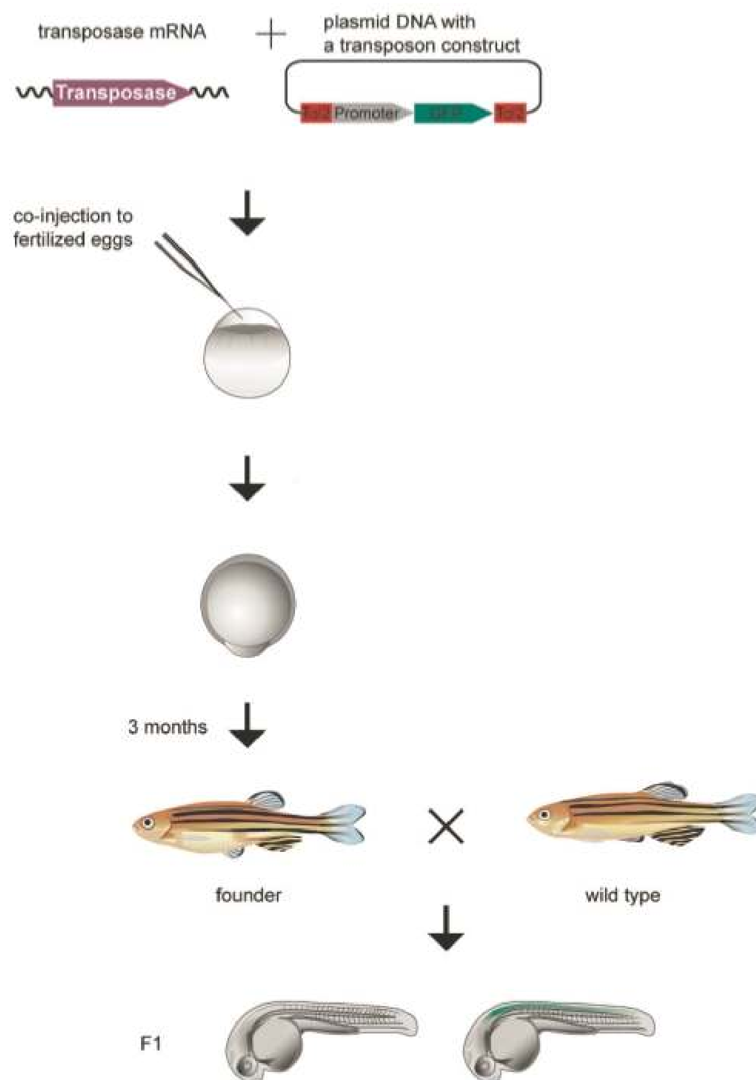


Figure 5. Tol2 tranposase technique. Schematic showing a co-injection, in one cell stage, of the mRNA transposase with the linearized construct in a specific vector. This will result in single-copy mode insertion in the genome. The fish that will achieve adulthood will be outcrossed and create the F1 generation. Adapted from [79].

The second technique consists on the injection of a pseudotype retrovirus in the blastula-stage embryos generating a chromosomal integration of its cDNA. The proviral insertion is transmitted to F1 generation and, although it becomes a germline transmitting founder fish with a 100% of efficacy, handling and modifying these retroviral vectors is more difficult when compared to Tol2 manipulation and construct generation [80, 81]. Because of this disavantage, the Tol2 tranposase technique was chosen.

3.2 **Microinjection**

Injections were performed using a Leica M165C stereomicroscope and a Pneumatic Picopump PV820 microinjection device (World Precision Instruments).
Injection volume was setup in mineral oil using the highest magnification objectives (40x) to achieve the highest precision.

An injection volume of 4 nL was used since smaller volume injection prevents malformations in contrast to a higher volume one. Eggs were injected in the yolk sac at the one cell stage. They were also injected with rhodamine to ensure that the injection process was correctly performed. Embryos were raised at 28°C. During the first hours after injections and in the following 5 days, the medium was replaced and the embryos were evaluated using a stereomicroscope: dead and/or unfertilised eggs were removed and surviving embryos were counted. Embryos expressing high fluorescence levels were kept growing before they were crossed with Wt zebrafish to investigate for potential founders (the founder fish crossed with Wt fish will generate a progeny with 10-15% green fluorescent embryos).

3.3 **Zebrafish culture and maintenance**

Zebrafish were maintained at 28°C on a 14 hour light / 10 hour dark cycle. They were housed in multi-tank flow through systems under standardised environmental conditions including continuous aeration, filtration, ultraviolet sterilization, and daily water replacement (10% of total circulating volume). Breedings were set up using a 1:1 male to female ratio. Embryos were collected during the morning and maintained in 10cm Petri dishes with a maximum of 80 embryos per dish. They were raised at 28°C in Petri dishes with embryo medium until they achieve 5 days of growth. At 5 days post-fertilization (dpf), the larvae were placed in 1 liter tanks filled with system water and fed with paramecia.

3.4 **Blasting**

The amino-acid sequences of human ATP13a2, Lrrk2, synphilin-1 and Glucocerebrosidade were blasted against the zebrafish genome sequence using Ensembl's website at www.ensembl.org [82].

3.5 **Phylogenetic tree and Functional analysis**

The blast highest scored sequences for each gene were chosen. To determine the phylogenetic proximity between each human and zebrafish sequences, protein sequences multiple alignment was performed with Clustal-X2 [83]. BioEdit software was used to visualize the phylogenetic tree and to determine the "identity" and "similarity" between the homologues sequences. Blosum62 matrix was used to obtain the similarity values [84].

Functional Analysis was determine with the software S.M.A.R.T provided by http://smart.embl-heidelberg.de/ [85] and the Reverse psi-blast program available at http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi [86].

To characterize both spatially and temporally the expression of ATP13a2 and Lrrk2:

3.6 Design of probes for *in-situ* hybridization

Forward and reverse primers for *ATP13a2* (CCTGACACTGAACCCCTCAT and CACTATCAAACAGGCCATGC, respectively) and LRRK2 (CCCTAAACCGCAGAGTATCA and ATTCATAGTCCACCGGTCTG) were chosen using the online version of the software Primer3 [87]. Total mRNA from 70 embryos with 24 hours was obtained using RNeasy mini Kit from Qiagen. The mRNA was transcribed to cDNA with Reverse Transciptase Polimerase Chain Reaction (RT-PCR) technique using Phusion enzyme from Finnyime. PCR was performed to get the specific 762bp ATP13a2 cDNA sequence. The fragment was transcribed in two different RNA probes with digoxigenine labeled nucleotides, using specific RNA polymerases (SP6 and T7 from Promega). These probes (anti-sense and sense) were further used in *in situ* experiments. The sense probe works has a control probe. Its sequence is identical to the sequence of the gene mRNA and therefore no interaction between the two is possible. The anti-sense probe has a complementary sequence to the mRNA of our gene so they will interact and form stable dsRNA.

No results for LRRK2 WISH will be presented since the sense probe (control) showed signaling indicating that it was hybridizing with some mRNA in the embryo. Therefore no conclusions could be taken and new primers were designed.

3.7 ***Whole mount In-situ*** **hybridization**

The in situ hybridization technique allows the expression pattern of a particular gene to be detected. In Zebrafish, this expression analysis can be done in the entire embryo due to its transparency in the early stages. Embryos with more than 24hpf were treated with 1-Phenyl-2-thiourea (PTU), a reagent that eliminates the natural pigmentation that starts to appear around the 23hpf.

The RNA probes with digoxigenine labeled nucleotides will dimerize with the mRNA that is present in the embryo forming stable dsRNA. An antibody for digoxigenine will recognizes these dsRNA and an enzymatic reaction will originate a purple color. Throw this purple color we are able to analyze the pattern of expression in the embryo.

Two sets of *in-situ* with the sense and anti-sense probes were done for ATP13a2 using embryos with 12, 24, 48, 72 hours and 6 days of life. This provides a temporal and spatial knowledge of the gene expression pattern during zebrafish embryonic development. The protocol was done accordingly to [88] .

3.8 **Epifluorescence Microscopy**

All images from tg Zebrafish and from *in situ* experiments were acquired using the Leica MZ10F microscope.

4. Results

Transgenic zebrafish lines for PD using human a-syn were initiated. To validate this model it will be required subsequent validation.

We also investigated four zebrafish orthologues of PD- associated genes using a bioinformatic approach, and found that the homology between the two species is high. In particular, the expression analysis of ATP13a2 revealed that the gene is expressed during embryonic development at different locations throughout the development which can reflect a differential role in neurogenesis.

Overall, these data point Zebrafish as an interesting model to study PD and other neurodegenerative disorders.

4.1 Generating the transgenic zebrafish lines

Three different DNA constructs were developed to generate three independent transgenic lines: THrat:a-syn:IRES:GFP, THrat:IRES:GFP and THzeb:GFP. The construct THrat:a-syn:IRES:GFP will generate the first stable transgenic line expressing the human a-syn. The other two constructs will generate two control lines expressing the EGFP under the control of rat TH (THrat) or zebrafish TH (THzeb), respectively. The work is part of a wider project that aims the development of several zebrafish transgenic lines of PD.

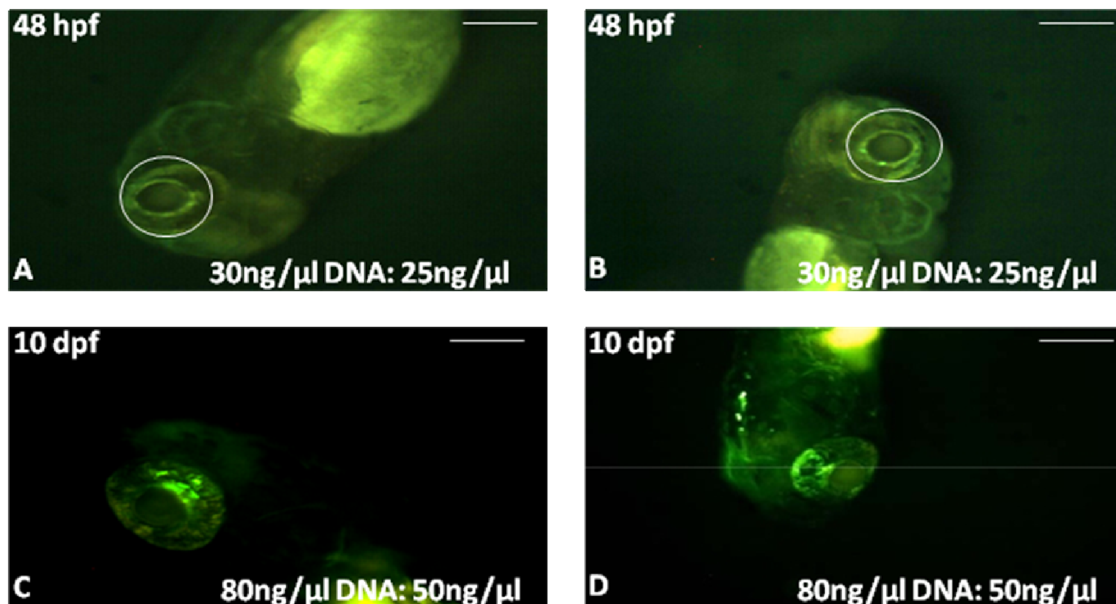
Those constructs were already microinjected in one cell stage embryos and they were checked for EGFP expression. Different concentrations of the constructs and mRNA transposase were initially tested with THrat:IRES:GFP. The lowest concentration resulting in protein fluorescence detection was 30ng/µl DNA: 25ng/µl mRNA transposase. However, as the intensity level of fluorescence at 48 hpf was very low (Figure 6 A-B), higher concentrations were investigated. The top concentration used to generate the founder fish was 80 ng/µl DNA: 50 ng/µl mRNA transposase. Higher DNA concentrations lead to an increased anatomical problems and higher lethality. With

the latter concentration, at 10 dpf, it is possible to detect protein fluorescence especially in the eye area (Figure 6 C-D). The expression pattern in adult brain will be analyzed in the F1 generation. For the THrat:a-syn:IRES:GFP, the same concentrations were investigated. It is possible to detect protein fluorescence at 48 hpf and 10 dpf. At 10 dpf the fluorescence is also localized in the retina (Figure 6 E-F).

In the THzeb:GFP construct, the DNA concentration ratio of 30 ng/µl:25 ng/µl mRNA transposase was enough to detect strong green fluorescence in all body of microinjected fish, designated as potential founders (Figure 6 G-H). All potential founders will grow until adulthood and then each one will be crossed with Wt zebrafish and positive, fluorescent, F1 generation is expected.

The protein expression pattern is more relevant in the fish brain and in the eye (Figure G-H).

No fluorescence is seen in control, non-injected, fish (Figure 6 I-J).



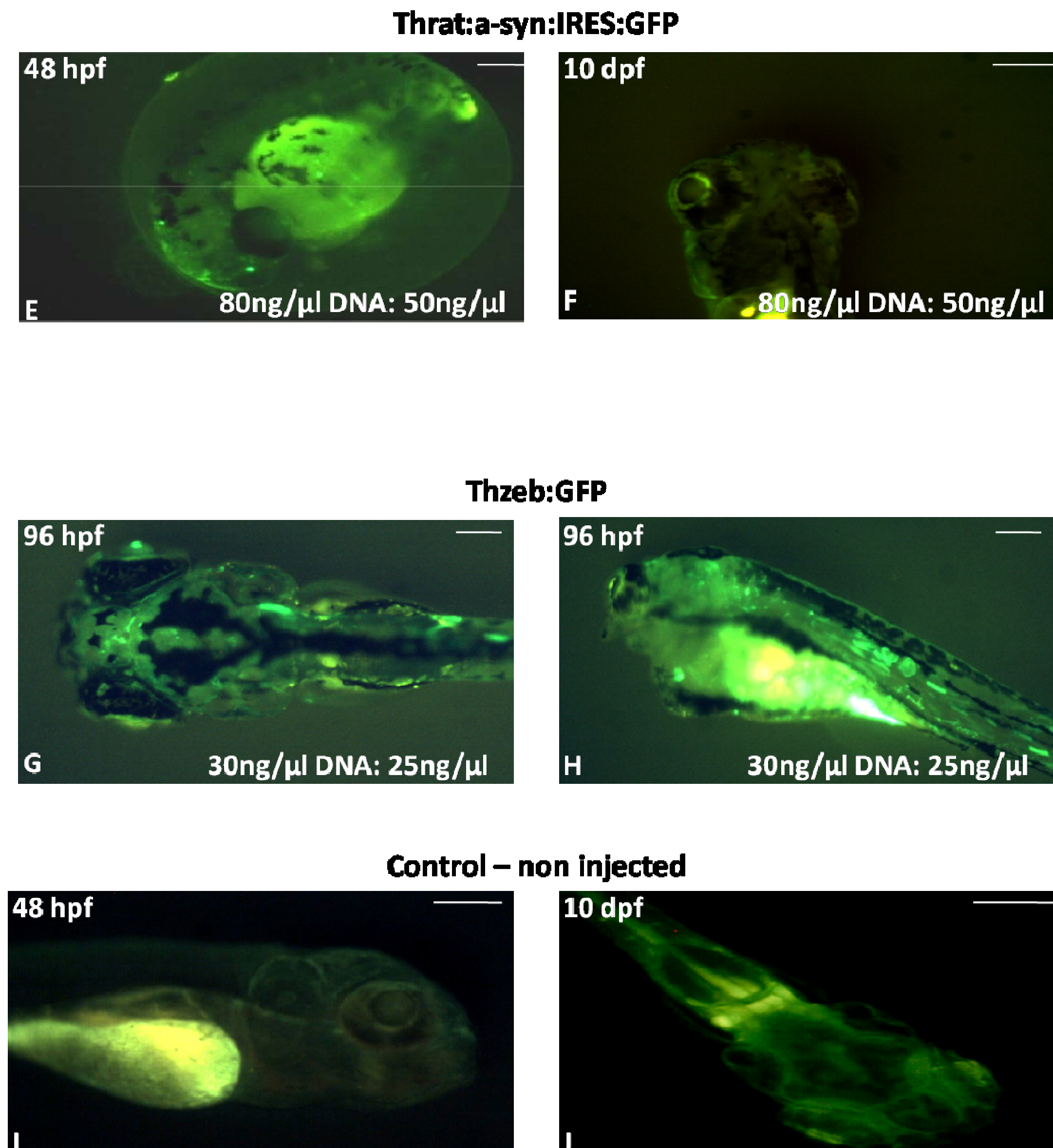


Figure 6. PD transgenic Zebrafish protein fluorescence. A-B) Low levels of fluorescence detectable with 30ng/µl DNA: 25ng/µl for the THrat:IRES:GFP construct injection at 48hpf. **C-D)** Founder Fishes protein fluorescence for THrat:IRES:GFP construct injection with 80ng/µl DNA: 50ng/µl mRNA transposase at 10dpf. **E-F)** Fishes injected with the construct THrat:a-syn:IRES:GFP at 48hpf and 10dpf. **G-H)** THzeb:GFP founder fishes with high protein fluorescence levels at 96hpf. **I-J)** Absence of protein fluorescence detection in non-injected zebrafish. Scale bar: 200µm

For the THzeb:EGFP construct, the potential founders fishes already reached adulthood and were crossed with wt fishes with a positive F1 generation being obtained. In this F1 the expression is localized in the brain (Figure 7 A-B)

Thzeb:EGFP F1 generation

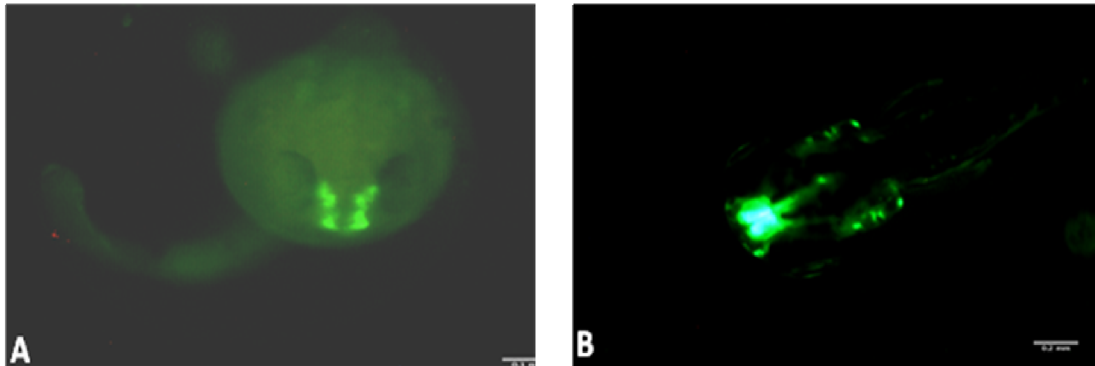


Figure 7. F1 positive Generation. The outcross of the founder fishes with the construct THzeb:EGFP revealed embryos with fluorescence. This fluorescence is focused in the brain area. Scale bar 200µm

4.2 **Bioinformatic analysis revealed high degree of conservation of PD-associated genes between Human and Zebrafish**

Since a-syn, 15 more genes were associated to PD [68-71]. From those, we focused our attention in the four genes linked to PD less explored: ATP13a2, Lrrk2, Synphilin-1 and Glucocerebrosidase. A search for a-syn has also performed but no sequence was found in zebrafish genome so far.

The blast the protein sequences of the four studied genes against Zebrafish genome revealed one high score match for each.

4.2.1 Zebrafish has one orthologue for Human ATP13a2-5

The human ATP13A family has five members, ATP13a1-5 (NCBI accession numbers NP_065143, NP_071372, NP_078800, NP_115655 and NP_940907 respectively).

Blasting the protein sequence of the human ATP13a2 against zebrafish genome identified the orthologue zgc:136762. Blasting the protein sequence from zebrafish against the human database showed that zgc:136762 was highly homologous to ATP13A2, -3, -4 and -5 with 50%, 37%, 38% and 35% of homology, respectively. Concordantly, the reverse comparison of those four human protein members sequences against the zebrafish genome, identified only zgc:136762 as their orthologue.

For human ATP13a1, the blast showed only one sequence: NP_001001403.1 with a homology between them of 74%.

Zebrafish has two orthologues for the five human ATP13ases protein. zgc:136762 is the orthologue of ATP13a2-5. This protein as a higher homology to the human ATP13a2 and therefore was named ATP13a2. The other zebrafish protein, ATP13a, is identical to ATP13a1 (Figure 8-9). These results were confirmed by phylogenetic analysis and they confirm that ATP13a2 and ATP13a1 share a higher homology with their zebrafish orthologues than with the other human ATP13ases (Figure 9).

```
ATP13a4_Hu    ---------- --MGHFEKGQ HALLNEGEE- -----NEMEI FGYRTQGCRK
ATP13a5_Hu    ---------- -MEENSKKDH RALLNQGEE- -----DELEV FGYRDHNVRK
ATP13a3_Hu    ---------- -----MDREE RKTINQGQE- -----DEMEI YGYNLSRWKL
ATP13a2_Hu    MSADS---SP LVGSTPTGYG TLTIGTSIDP LSSSVSSVRL SGYCGSPWRV
zgc_136762    MDAKGNCVPG LDGASSSPDT EPLIKDPRPP ELSLVSHMDV QGYRWVCWKV
Clustal Co                             :            . : :  **     :

ATP13a4_Hu    SLCLAGSIFS FGILPLVFYW RPAWHVWAHC VPCSLQEADT VLLRTTDEFQ
ATP13a5_Hu    AFCLVASVLT CGGLLLVFYW RPQWRVWANC IPCPLQEADT VLLRTTDEFQ
ATP13a3_Hu    AIVSLGVICS GGFLLLLLYW MPEWRVKATC VRAAIKDCEV VLLRTTDEFK
ATP13a2_Hu    IGYHVVVWMM AGIPLLLFRW KPLWGVRLRL RPCNLAHAET LVIEIRDKED
zgc_136762    WLCRIGAVFS VGLLLVLFKW RPRIGILARC KSCPISMADV LLLKDRYG--
Clustal Co                *   ::: *  *   :       . :  .:. :::.

ATP13a4_Hu    IYSWKKVIWI YLSALNSAFG LTPDHPLMT- ---------- -DEEYIINRA
ATP13a5_Hu    RYMRKKVFCL YLSTLKFPVS KKWEESLVA- ---------- -DRHSVINQA
ATP13a3_Hu    MWFCAKIRVL SLETYPVSSP KSMSNKLSNG HAVCLIENPT EENRHRISKY
ATP13a2_Hu    SSWQLFTVQV QTEAIGEGSL EPSPQSQAED GRSQAAVGAV PEGAWKDTAQ
zgc_136762    ---QQFVVDV ITEEVEEGSL D--------- ----FAVGDA DENEWRDTVQ
Clustal Co             :   .                               :     .

ATP13a4_Hu    IRKP------ --DLKVRCIK VQKIRYVWNY LEGQFQKIGS LEDWLSSAKI
ATP13a5_Hu    LIKP------ --ELKLRCME VQKIRYVWND LEKRFQKVGL LEDSNSCSDI
ATP13a3_Hu    SQTE------ --SQQIRYFT HHSVKYFWND TIHNFDFLKG LDEGVSCTSI
ATP13a2_Hu    LHKSEEAVSV GQKRVLRYYL FQGQRYIWIE TQQAFYQVSL LDHGRSCDDV
zgc_136762    LHSE------ -KKTLLRYYV FEGIRYIWIS KKGAFCKASV LSEGWTCADL
Clustal Co      .          .  :*     .  :*.*       *      *..  :. .:
```

```
ATP13a4_Hu   HQKFGSGLTR EEQEIRRLIC GPNTIDVEVT PIWKLLIKEV LNPFYIFQLF
ATP13a5_Hu   HQTFGLGLTS EEQEVRRLVC GPNAIEVEIQ PIWKLLVKQV LNPFYVFQAF
ATP13a3_Hu   YEKHSAGLTK GMHAYRKLLY GVNEIAVKVP SVFKLLIKEV LNPFYIFQLF
ATP13a2_Hu   HRSR-HGLSL QDQMVRKAIY GPNVISIPVK SYPQLLVDEA LNPYYGFQAF
zgc_136762   HGQQ-QGLSR ADQSTRKQIF GANIIDVPVK SYLQLLFEEV LNPFYIFQVF
Clustal Co   :     **:    :  *: :  * * * : :  .  :**..:. ***:* ** *

ATP13a4_Hu   SVCLWFSEDY KEYAFAIIIM SIISISLTVY DLREQSVKLH HLVESHNSIT
ATP13a5_Hu   TLTLWLSQGY IEYSVAIIIL TVISIVLSVY DLRQQSVKLH NLVEDHNKVQ
ATP13a3_Hu   SVILWSTDEY YYYALAIVVM SIVSIVSSLY SIRKQYVMLH DMVATHSTVR
ATP13a2_Hu   SIALWLADHY YWYALCIFLI SSISICLSLY KTRKQSQTLR DMVKLS--MR
zgc_136762   SIILWMSDGYf VYYAACIFII SLISIGVSLY ETRKQSTTLR RMACLI--VN
Clustal Co   :: ** :: *   *: .*.:: : :**  ::* . *:*   *:  :.     :

ATP13a4_Hu   VSVCGRKAGV QELESRVLVP GDLLILTGNK VLMPCDAVLI EGSCVVDEGM
ATP13a5_Hu   VTIIVKDKGL EELESRLLVP GDILILPG-K FSLPCDAVLI DGSCVVNEGM
ATP13a3_Hu   VSVCRVNEEI EEIFSTDLVP GDVMVIPLNG TIMPCDAVLI NGTCIVNESM
ATP13a2_Hu   VCVCRPGGEE EWVDSSELVP GDCLVLPQEG GLMPCDAALV AGECMVNESS
zgc_136762   VTVRRDTGEE ECVSSEELVP GDCVVIPAEG LLLPCDAALV AGECMVNESM
Clustal Co   * :        : : *  *** ** :::.      :****.*:  * *:*:*.

ATP13a4_Hu   LTGESIPVTK TPLPKMDSSV P---WKTQSE ADYKRHVLFC GTEVIQAKAA
ATP13a5_Hu   LTGESIPVTK TPLPQMENTM P---WKCHSL EDYRKHVLFC GTEVIQVKPS
ATP13a3_Hu   LTGESVPVTK TNLPNPSVDV KGIGDELYNP ETHKRHTLFC GTTVIQTRFY
ATP13a2_Hu   LTGESIPVLK TALPEG---- ----LGPYCA ETHRRHTLFC GTLILQAR-A
zgc_136762   LTGESIPVMK TPLSNS---- ----EATYNP ESQRRHTLFC GTQIIQAKGG
Clustal Co   *****:** * * *.:                    ::*.*** ** ::*.:

ATP13a4_Hu   CSGT-VRAVV LQTGFNTAKG DLVRSILYPK PVNFQLYRDA IRFLLCLVGT
ATP13a5_Hu   GQGP-VRAVV LQTGYNTAKG DLVRSILYPR PLNFKLYSDA FKFIVFLACL
ATP13a3_Hu   TGEL-VKAIV VRTGFSTSKG QLVRSILYPK PTDFKLYRDA YLFLLCLVAV
ATP13a2_Hu   YVGPHVLAVV TRTGFCTAKG GLVSSILHPR PINFKFYKHS MKFVAALSVL
zgc_136762   GPGKGAIAVV TCTGFLTAKG DLISSILYPQ PLDFRFYRDA MKFLLFLGLL
Clustal Co        . *:*   **: *:**  *: ***:*: * :*::* .:   *:  *

ATP13a4_Hu   ATIGMIYTLC VYVLSGEPPE EVVRKALDVI TIAVPPALPA ALTTGIIYAQ
ATP13a5_Hu   GVMGFFYALG VYMYHGVPPK DTVTMALILL TVTVPPVLPA ALTIGNVYAQ
ATP13a3_Hu   AGIGFIYTII NSILNEVQVG VIIIESLDII TITVPPALPA AMTAGIVYAQ
ATP13a2_Hu   ALLGTIYSIF ILYRNRVPLN EIVIRALDLV TVVVPPALPA AMTVCTLYAQ
zgc_136762   ALIGTIYSLV ILSKSNTPWK ELIIRSLDIV TIIVPPALPA AITTATIYAQ
Clustal Co   . :* :*::               :  :* :: *: ***.*** *:*   :***

ATP13a4_Hu   RRLKKRGIFC ISPQRINVCG QLNLVCFDKT GTLTRDGLDL WGVVSCDRNG
ATP13a5_Hu   KRLKKKKIFC ISPQRINMCG QINLVCFDKT GTLTEDGLDL WGTVPTADNC
ATP13a3_Hu   RRLKKIGIFC ISPQRINICG QLNLVCFDKT GTLTEDGLDL WGIQRVENAR
ATP13a2_Hu   SRLRRQGIFC IHPLRINLGG KLQLVCFDKT GTLTEDGLDV MGVVPLKGQA
zgc_136762   NRLKRQGVFC ISPPRINICG KISLFCFDKT GTLTEEGLDV WGVMEVTGGV
Clustal Co    **::  :** * * ***: * ::.*.***** ****.:***:  *

ATP13a4_Hu   FQEVHSFASG QALPWGPLCA AMASCHSLIL LDGTIQGDPL DLKMFEATTW
ATP13a5_Hu   FQEAHSFASG QAVPWSPLCA AMASCHSLIL LNGTIQGDPL DLKMFEGTAW
ATP13a3_Hu   FLSPEENVCN EMLVKSQFVA CMATCHSLTK IEGVLSGDPL DLKMFEAIGW
ATP13a2_Hu   FLPLVP--EP RRLPVGPLLR ALATCHALSR LQDTPVGDPM DLKMVESTGW
zgc_136762   FGELVP--DP LFLPPGLMLS ALASCHSVAL LGGQALGDPL ELKMIESTGW
Clustal Co   *            :  . :   .:*:**::   : .   ***: :***.*.  *

ATP13a4_Hu   EMAFS---GD DFHIKGVPAH AMVVKPCRTA SQVPVEG--- ----------
ATP13a5_Hu   KMEDC---IV DSCKFGT-SV SNIIKPGPKA SKSPVEA--- ----------
ATP13a3_Hu   ILEEA---TE EETALHNRIM PTVVRPPKQL LPESTPAGNQ EMELFELPAT
ATP13a2_Hu   VLEE---EPA ADSAFG-TQV LAVMRPPLWE PQLQAMEE-- --------PP
```

```
zgc_136762   ELTEPENDMG HDSEFGGHRV LAVMRPPASE LLTEGNS--- --------VS
Clustal Co    :                      :::*

ATP13a4_Hu   --IAILHQFP FSSALQRMTV IVQEMGGDRL -AFMKGAPER VASFCQPETV
ATP13a5_Hu   --IITLCQFP FSSSLQRMSV IAQLAGENHF HVYMKGAPEM VARFCRSETV
ATP13a3_Hu   YEIGIVRQFP FSSALQRMSV VARVLGDRKM DAYMKGAPEA IAGLCKPETV
ATP13a2_Hu   VPVSVLHRFP FSSALQRMSV VVAWPGATQP EAYVKGSPEL VAGLCNPETV
zgc_136762   QPVAIVRRFP FSSSLQRMSV VTVGPAEPSP VAFIKGAPEM VASFCHKESV
Clustal Co     :  : :** ***:****:* :.   .      .::**:**  :* :*. *:*

ATP13a4_Hu   PTSFVSELQI YTTQGFRVIA LAYK--KLEN ---DHHATTL TRETVESDLI
ATP13a5_Hu   PKNFPQELRS YTVQGFRVIA LAHKTLKMGN ---LSEVEHL AREKVESELT
ATP13a3_Hu   PVDFQNVLED FTKQGFRVIA LAHR--KLES KLTWHKVQNI SRDAIENNMD
ATP13a2_Hu   PTDFAQMLQS YTAAGYRVVA LASK--PLPT VPSLEAAQQL TRDTVEGDLS
zgc_136762   PSHFSHTLRE YASQGFRVLG LAYK--HLAK ETDLSTVERV E---VEKGMN
Clustal Co   *  *   *.  ::  *:**:. ** :   : .       .  :     :*  :

ATP13a4_Hu   FLGLLILENR LKEETKPVLE ELISARIRTV MITGDNLQTA ITVARKSGMV
ATP13a5_Hu   FLGLLIMENR LKKETKLVLK ELSEARIRTV MITGDNLQTA ITVAKNSEMI
ATP13a3_Hu   FMGLIIMQNK LKQETPAVLE DLHKANIRTV MVTGDSMLTA VSVARDCGMI
ATP13a2_Hu   LLGLLVMRNL LKPQTTPVIQ ALRRTRIRAV MVTGDNLQTA VTVARGCGMV
zgc_136762   FLGLLVMKNQ VKPESAEVIQ TLTLAQLRPV MVTGDNILTA VNVARVCGMV
Clustal Co   ::**:::.*  :* ::  *::  *  :.:*.* *:***.: ** :.**: . *:

ATP13a4_Hu   SESQKVILIE ANETTGSSSA SISWTLVEEK KHIMYG---N QDNYINIRDE
ATP13a5_Hu   PPGSQVIIVE ADEPEEFVPA SVTWQLVE-N QETGPG---K KEIYMHTGNS
ATP13a3_Hu   LPQDKVIIAE ALPPKDGKVA KINWHYADSL TQCSHPSAID PEAIPVKLVH
ATP13a2_Hu   APQEHLIIVH ATHPERGQPA SLEFLPMESP TAVNGVK--D PDQAASYTVE
zgc_136762   PLHEKVIFVH ASPPTAGSQA SLQFHEGEGA TATINTQ--Q TIDIPVQGQY
Clustal Co      .::*: . *  .     * .: :   :            .

ATP13a4_Hu   VSDKGREGS- YHFALTGKSF HVISQHFSSL LPKILINGTI FARMSPGQKS
ATP13a5_Hu   STPRGEGGSC YHFAMSGKSY QVIFQHFNSL LPKILVNGTV FARMSPGQKS
ATP13a3_Hu   DSLEDLQMTR YHFAMNGKSF SVILEHFQDL VPKLMLHGTV FARMAPDQKT
ATP13a2_Hu   PDPRS----- RHLALSGPTF GIIVKHFPKL LPKVLVQGTV FARMAPEQKT
zgc_136762   QSAVG----- YHLAINGMSF AALCDHFPEY LPKVLMRGTI YARMTPEQKT
Clustal Co       .       *:*:.* ::   : .**  .  :**:::.**: :***:* **:

ATP13a4_Hu   SLVEEFQKLD YFVGMCGDGA NDCGALKMAH VGISLSEQEA SVASPFTSKT
ATP13a5_Hu   SLIEEFQKLN YYVGMCGDGA NDCGALKAAH AGISLSEQEA SVASPFTSKT
ATP13a3_Hu   QLIEALQNVD YFVGMCGDGA NDCGALKRAH GGISLSELEA SVASPFTSKT
ATP13a2_Hu   ELVCELQKLQ YCVGMCGDGA NDCGALKAAD VGISLSQAEA SVVSPFTSSM
zgc_136762   QLVKALQKLN YRVGMCGDGA NDCGALRAAD VGVSLSDAEA SVASPFTSKS
Clustal Co   .*:  :*::: * ******** ******: *.  *:***: ** **.*****.

ATP13a4_Hu   PNIECVPHLI KEGRAALVTS FCMFKYMALY SMIQYVGVLL LYWETNSLSN
ATP13a5_Hu   TNIQCVPHLI REGRAALVSS FGVFKYLTMY GIIQFISALL LYWQLQLFGN
ATP13a3_Hu   PSISCVPNLI REGRAALITS FCVFKFMALY SIIQYFSVTL LYSILSNLGD
ATP13a2_Hu   ASIECVPMVI REGRCSLDTS FSVFKYMALY SLTQFISVLI LYTINTNLGD
zgc_136762   DNISCVPLLI KEGRCSLVTS FSLFKYMALY SLIQFASVLI LYTEKTNLGD
Clustal Co    .*.*** :* :***.:* :* * :**::::* .: *: .. : **     :.:

ATP13a4_Hu   YQFLFQDLAI TTLIGVTMNL NGAYPKLVPF RPAGRLISPP LLLSVIFNIL
ATP13a5_Hu   YQYLMQDVAI TLMVCLTMSS THAYPKLAPY RPAGQLLSPP LLLSIFLNSC
ATP13a3_Hu   FQFLFIDLAI ILVVVFTMSL NPAWKELVAQ RPPSGLISGA LLFSVLSQII
ATP13a2_Hu   LQFLAIDLVI TTTVAVLMSR TGPALVLGRV RPPGALLSVP VLSSLLLQMV
zgc_136762   LQFLFFDLVL VTVLAILMGR GGPSNDLHPQ RPAASLLSLP VLASLLMHTV
Clustal Co    *:*  *:.:    : . *.    .   *    **.. *:* . :* *:: :

ATP13a4_Hu   LSLAMHIAGF ILVQRQPWYS -VEIHSACTV QNESISELTM SPTAPEKMES
ATP13a5_Hu   FSCIVQISAF LYVKQQPWYC EVYQYSECFL ANQ--SNFST NVSLERNWTG
```

```
ATP13a3_Hu   ICIGFQSLGF FWVKQQPWYE VWHPKSDACN TTG--SGFWN SSHVDNETEL
ATP13a2_Hu   LVTGVQLGGY FLTLAQPWFV PLNRTVAAPD NLP------- ----------
zgc_136762   LLILAQVSGL LITMSQDWYV PLNSTRTGAA NLP------- ----------
Clustal Co   :    :  .  : .  * *:

ATP13a4_Hu   NSTFT----- SFENTTVWFL GTINCITVAL VFSKGKPFRQ PTYTNYIFVL
ATP13a5_Hu   NATLIPGSIL SFETTTLWPI TTINYITVAF IFSKGKPFRK PIYTNYIFSF
ATP13a3_Hu   DEHNIQ---- NYENTTVFFI SSFQYLIVAI AFSKGKPFRQ PCYKNYFFVF
ATP13a2_Hu   ---------- NYENTVVFSL SSFQYLILAA AVSKGAPFRR PLYTNVPFLV
zgc_136762   ---------- NMEDTSVFAV SGFQYIIMSV VITKGFPYKK PLYHNVLFVC
Clustal Co              . * * :: :   :: : ::   .:** *::: * * *  *

ATP13a4_Hu   VLIIQLGVCL FILFADIPEL YRRLDLLCTP VL-WRASIVI MLSLNFIVSL
ATP13a5_Hu   LLLAALGLTI FILFSDFQVI YRGMELIPTI TS-WRVLILV VALTQFCVAF
ATP13a3_Hu   SVIFLYIFIL FIMLYPVASV DQVLQIVCVP YQ-WRVTMLI IVLVNAFVSI
ATP13a2_Hu   ALALLSSVLV GLVLVPG-LL QGPLALRNIT DTGFKLLLLG LVTLNFVGAF
zgc_136762   ALVFLFALMS WLVLFRHTII HRVLSLYDIT DMSYKLLLVA IAALNFFICF
Clustal Co    :     .    :::     :    : :        ::  ::  :   :   .:

ATP13a4_Hu   VAEEAVIENR ALWMMIKRCF GYQS-KSQYR IWQRDLANDP SWPPLNQTSH
ATP13a5_Hu   FVEDSILQNH ELWLLIKREF GFYS-KSQYR TWQKKLAEDS TWPPINRTDY
ATP13a3_Hu   TVEESVDRWG KCCLPWALGC RKKTPKAKYM YLAQELLVDP EWPPKPQTTT
ATP13a2_Hu   MLESVLDQCL PACLRRLR-- PKRASKKRFK QLERELAEQP -WPPLPAGPL
zgc_136762   LLEFFIDQGA LNCLRNLR-- GKRESKKQYK RLNVQLTETP SWPPLNQPLF
Clustal Co     *  : .      :            *  ::      .*   .  ***

ATP13a4_Hu   SDMPECGRGV SYSNPVFESN EEQL------ ---------- ---
ATP13a5_Hu   SGD---GKNG FYINGGYESH EQIPKRKLKL GGQPTEQHFW ARL
ATP13a3_Hu   EAK-----AL VKENGSCQII TIT------- ---------- ---
ATP13a2_Hu   R--------- ---------- ---------- ---------- ---
zgc_136762   PSQ------- ----SSVISI S--------- ---------- ---
Clustal Co
```

Figure 8. Multiple protein sequence alignment of human ATP13a2-5 and the orthologue zgc_136762 in Zebrafish. The alignment was performed by ClustalX2 ("*" identical residues,":" conserved substitutions,"." semi-conserved substitutions). The yellow background highlights the E1_E2 ATPase domain and the green one the HAD_like domain.

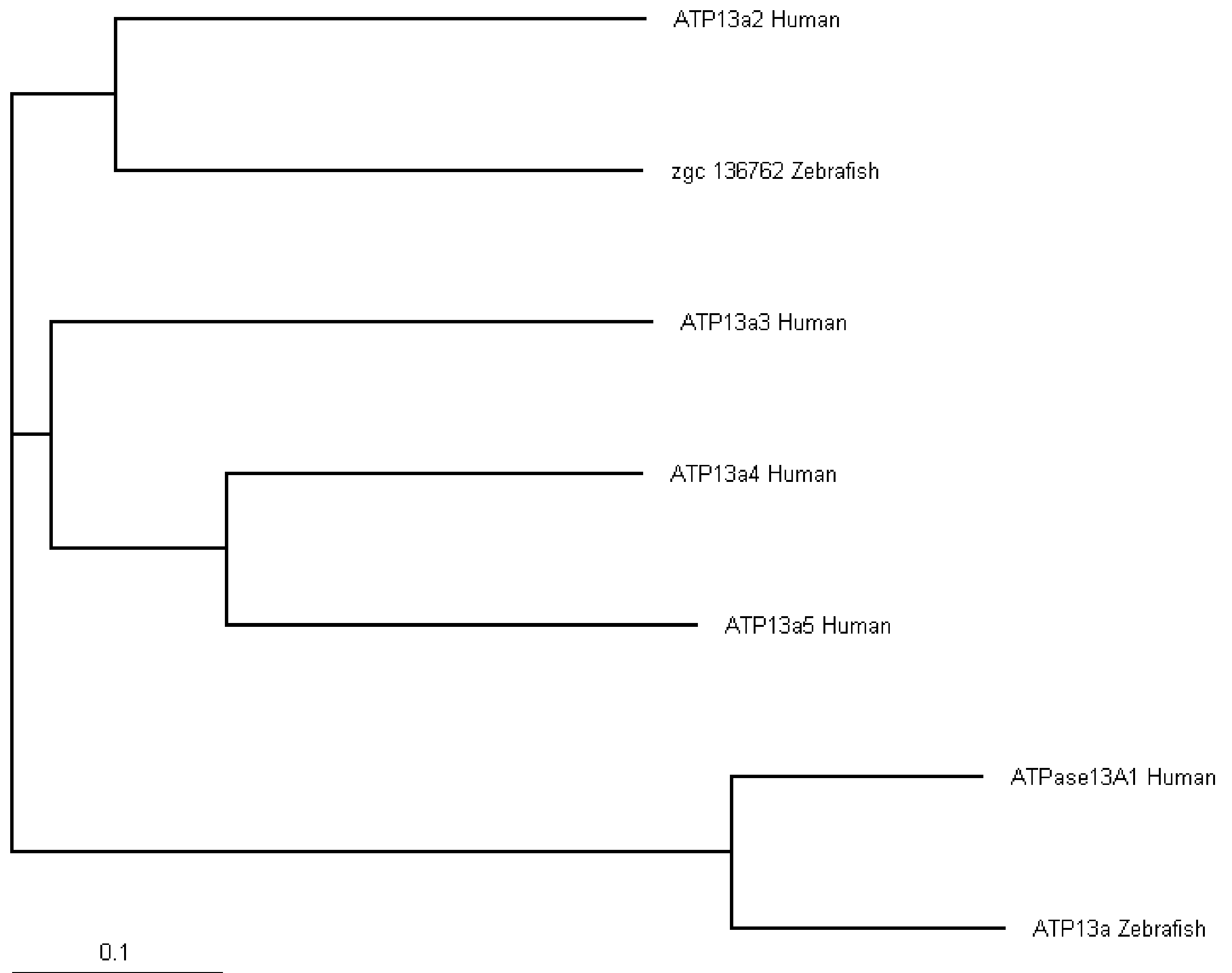


Figure 9. Phylogenetic tree including the human ATP13a -2 to -5 and the orthologue in Zebrafish. The zebrafish orthologue is closer to the human ATP13a2 than to the other ATP13a3-5. ATP13a1 is a separate case, being closer to a different protein in Zebrafish named ATP13a.

The human and Zebrafish ATP13a2 proteins are very similar in size with 1180 and 1170 amino-acids, respectively. Phylogenetic analysis revealed that the proteins share 50% Identity and 69% similarity in the total sequence. Analysis of the conserved domains showed that the identity percentage is maintained around 50% (Figure 10). The E1-E2 ATPase can be found in several species from bacteria to eukaryotic plasma and organelles. It is involved in different ions transportation across membranes using ATP hydrolysis for energy [89]. The haloacid dehalogenase (HAD_like) domain is found in proteins able to catalyze carbon or phosphoryl group transfer on several types of substrate, through a nucleophilic attack by an invariant active site aspartate [90].

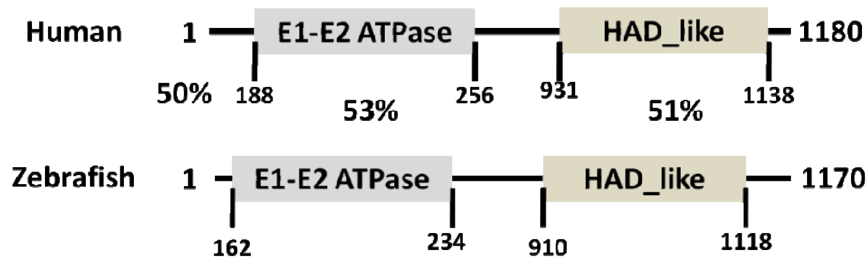


Figure 10. Side by side comparison of Human and Zebrafish ATP13a2. Both sequences share a similar size, 50 % homology and 69% of similarity. The conserved functional domains E1-E2 ATPase and HAD_like have approximately the same identity (52% and 51%, respectively) and similarity (66% and 69%, respectively).

4.2.2 Human and Zebrafish Lrrk2 display 66% similarity

Concerning the human Lrrk2 protein (NP_940980.3), the blast revealed one matching sequence, ADJ67257. The two proteins have a similar size, 2533 bp and 2527 bp in zebrafish and human respectively. The alignment of the two sequences showed a 47% identity and 66% homology (Figures 11). All the functional domains are conserved between the species but with a very different homology rates, ranging from 10% to 64% (Figure 11 and 12).

```
D. rerio      -----MAEIE ELSIRLKKLL VRLN-LQDGK QLGTMVQIIE DLLFLSHTEH
H. sapiens    MASGSCQGCE EDEETLKKLI VRLNNVQEGK QIETLVQILE DLLVFTYSEH
Alignment              * * .  ****: **** :*:** *: *:***:* ***.::::**

D. rerio      CVELFADQNV HVPVLLKCSE SASIK-VQQV GWSLLCRLME ICPNTLDNLA
H. sapiens    ASKLFQGKNI HVPLLIVLDS YMRVASVQQV GWSLLCKLIE VCPGTMQSLM
Alignment     . :** .:*: ***:*:  ..     :  **** ******:*:* :**.*::.*

D. rerio      RP----MDYE FIDAHKQILK VLHEYHNKDA KMMMVALRAL ALMLKSGEIK
H. sapiens    GPQDVGNDWE VLGVHQLILK MLTVHN-ASV NLSVIGLKTL DLLLTSGKIT
Alignment      *     *:* .:..*: *** :*  ::  .. :: ::.*::*  *:*.**:*.

D. rerio      MQVLDEEEWD VFYSILEAMK SFSDREEVQL QGCTALQPLL QTVSEYHLAE
H. sapiens    LLILDEE-SD IFMLIFDAMH SFPANDEVQK LGCKALHVLF ERVSEEQLTE
Alignment     : :****  * :*  *::**: **. .:***  **.**: *: : *** :*:*


D. rerio      FIEKKDHEVV LNALGCFMDS ENVVLEALKV LIPLANPASN VETLMSKTVK
H. sapiens    FVENKDYMIL LSALTNFKDE EEIVLHVLHC LHSLAIPCNN VEVLMSGNVR
Alignment     *:*:**: :: *.**  * *. *::**..*:  * .** *..* **.*** .*:
```

```
D. rerio     CHSLTCRAMN TWLDSEAIQE AGCCLLWKFT SKGYYDMLVL NGVHKVAVKA
H. sapiens   CYNIVVEAMK AFPMSERIQE VSCCLLHRLT LGNFFNILVL NEVHEFVVKA
Alignment    *:.:. .**: ::  ** *** ..**** ::*   .::::*** * **:..***

D. rerio     CVSYPDNAIL QTAALSCLSA LAECIVQNGG LDEEWNEEDE EEQKVLVKKE
H. sapiens   VQQYPENAAL QISALSCLAL LTETIFLNQD LEEKNENQEN DDE-------
Alignment      .**:** * * :*****: *:* *. * . *:*: ::::: :::

D. rerio     AAQAEEMLIW REACYTAFER HAEDVKVQEA ACWTLNSLLL HCN--TSNHV
H. sapiens   --GEEDKLFW LEACYKALTW HRKNKHVQEA ACWALNNLLM YQNSLHEKIG
Alignment        *: *:*  ****.*:   * :: :**** ***:**.**: : *    .:

D. rerio     ELEGRPPLHT LIMAAMLLHS SSVKVFQAAS STLRTLIQRH CRIRSPLLSN
H. sapiens   DEDGHFPAHR EVMLSMLMHS SSKEVFQASA NALSTLLEQN VNFRKILLSK
Alignment    : :*: * *   :* :**:** ** :****:: .:* **::::  .:*. ***:

D. rerio     GIHFNIVELM RKHPNSSAVC ESACKLIHTL FQGARASLDD GFLILSQILI
H. sapiens   GIHLNVLELM QKHIHSPEVA ESGCKMLNHL FEGSNTSLDI MAAVVPKILT
Alignment    ***:*::*** :** :*. *. **.**::: * *:*:.:***      ::.:**

D. rerio     ALKTHTFLPE VQLEGLRASL VLLNPDRSLR EHGVSVADPD MVDVSLQVLK
H. sapiens   VMKRHETSLP VQLEALRAIL HFIVPGMPEE SR-----EDT EFHHKLNMVK
Alignment    .:* *      ****.*** *  :: *. . . .:      :    .. .*:::*

D. rerio     NQCVLEGAHT VYLQALNRFI SSESIQECGL GVLAALADSS GAVDLMCQQG
H. sapiens   KQCFKNDIHK LVLAALNRFI GNPGIQKCGL KVISSIVHFP DALEMLSLEG
Alignment    :**. :. *. : * ****** .. .**:***  *::::.. . .*::::. :*

D. rerio     AIDTVLHTLQ MFPQERDIHY WGLSLLFHLI SKKKLSRMMV PVLASVLVSS
H. sapiens   AMDSVLHTLQ MYPDDQEIQC LGLSLIGYLI TKKNVFIGTG HLLAKILVSS
Alignment    *:*:****** *:*::::*:   ****: :** :**::          :**.:****

D. rerio     VRKHKEDSVM LLKGLQVVWK LLDTCSSAAV WLQKEAFEKE IFQILRENTA
H. sapiens   LYRFKDVAEI QTKGFQTILA ILKLSASFSK LLVHHSFDLV IFHQMSSNIM
Alignment    : :.*: : :   **:*.:   :*. .:* :   * :.:*:   **: : .*

D. rerio     DQRRDPLQGM SCLCLSKMVM DGEILYALLE RACEDGDVDM AECLIQLGAD
H. sapiens   EQKDQQFLNL CCKCFAKVAM DDYLKNVMLE RACDQNNSIM VECLLLLGAD
Alignment    :*: : : .: .* *::*:.* *. :   .:** ***::.:  * .***: ****

D. rerio     VNKKTKSDSL LYQACDRGAP LSLLELLVSS GVHEQHLRGA LSVCVRRSDD
H. sapiens   ANQAKEGSSL ICQVCEKESS PKLVELLLNS GSREQDVRKA LTISIGKGDS
Alignment    .*: .:..** : *.*:: :.  .*:***:.* * :**.:* * *::.: :.*.

D. rerio     PAVTLILRRL GLDHTNSALC LGSVRIGHMK ASWISALLSE RRSQSTNTHY
H. sapiens   QIISLLLRRL ALDVANNSIC LGGFCIGKVE PSWLGPLFPD KTSNLR--KQ
Alignment      ::*:**** .** :*.::* **.. **::: .**:..*:.: : *:     :

D. rerio     SSKGQRLARQ ISQLQRKKGI VGVSRLSSDV STSGYFTDEE SDDSHIS--L
H. sapiens   TNIASTLARM VIRYQMKSAV EEGTASGSDG NFSEDVLSKF DEWTFIPDSS
Alignment    :. .. ***  : : * *..:     :  .**  . *  . .:  .: :.*.

D. rerio     EDSLVFMFDE LESDGSDGPP HGFMLISDSP EVSRRPVWRR RSHSRRTSSE
H. sapiens   MDSVFAQSDD LDSEGSEGS- --FLVKKKSN SISVGEFYRD AVLQRCSPNL
Alignment     **:.   *: *:*:**:*.   *:: ..* .:*   .:*     .* :..

D. rerio     GYHG------ --EADPSVPV QKHANSHNSR GQG------- FSESFSSPVV
H. sapiens   QRHSNSLGPI FDHEDLLKRK RKILSSDDSL RSSKLQSHMR HSDSISS-LA
Alignment      *.         . *    :*  .*.:*   ..         .*:*:** :.

D. rerio     LDKDPVRLLD LSGNELNDLS CLTDLNSLKK PIENLHRLDL SGNNLSQFPS
```

```
H. sapiens    SEREYITSLD LSANELRDID ALSQKCCISV HLEHLEKLEL HQNALTSFPQ
Alignment      ::: :  ** **.***.*:. .*::  .:.   :*:*.:*:*   * *:.**.

D. rerio      ILCQSLRSLT RLDLQGNHLQ CLPSELLSLP ALHTLNVSRN CIGPLLQLEP
H. sapiens    QLCETLKSLT HLDLHSNKFT SFPSYLLKMS CIANLDVSRN DIGPSVVLDP
Alignment      **::*:*** :***:.*::   .:** **.:. .: .*:****  *** : *:*

D. rerio      GVCSPALRRL NLSFNQITVC PFQLRSATQR LEELSLEGNQ ISELSLPLCL
H. sapiens    TVKCPTLKQF NLSYNQLSFV PENLTDVVEK LEQLILEGNK ISGICSPLRL
Alignment      *  .*:*::: ***:**::.   * :* ...:: **:* ****: **  :.  ** *

D. rerio      AELKVLDVSK NQVKIVSDNF LAECLKMETF IASVNQISSL PHLPSKITTV
H. sapiens    KELKILNLSK NHISSLSENF LEACPKVESF SARMNFLAAM PFLPPSMTIL
Alignment      ***:*::** *::.  :*:** *   * *:*:*  * :* :::: *.**..:* :

D. rerio      KLSHNTFTSV PEIVINLPCL RSVDMRNNSV GVLPGPSVWL SVNLRELMFS
H. sapiens    KLSQNKFSCI PEAILNLPHL RSLDMSSNDI QYLPGPAHWK SLNLRELLFS
Alignment     ***:*.*:.: **  ::*** * **:** .*.:    ****: *   *:*****:**

D. rerio      HNLISALDLS GPVYKWARLE KLHLSFNRLT EIPPQIGMLE DLTSLDVSHN
H. sapiens    HNQISILDLS EKAYLWSRVE KLHLSHNKLK EIPPEIGCLE NLTSLDVSYN
Alignment     ** ** ****    .* *:*:* *****.*:*. ****:** **  :*******:*

D. rerio      EGLRSFPDEM GKLVHLWDLP LDGLQLQLDL KHIGSKTKDI IRFLQQRLKK
H. sapiens    LELRSFPNEM GKLSKIWDLP LDELHLNFDF KHIGCKAKDI IRFLQQRLKK
Alignment       *****:** ***  ::**** ** *:*::*: ****.*:*** **********

D. rerio      AVPYHRMKLM VLGGTGSGKS SLIQQLMRLR RS---QWRSD PGVSIRDWPV
H. sapiens    AVPYNRMKLM IVGNTGSGKT TLLQQLMKTK KSDLGMQSAT VGIDVKDWPI
Alignment     ****:***** ::*.*****: :*:****: : :*       :    *:.::***:

D. rerio      RSKDKR--NM MLNVWEFSGG EECSGIHPHF MSSRAVYLVL YDLSKGASEI
H. sapiens    QIRDKRKRDL VLNVWDFAGR EEFYSTHPHF MTQRALYLAV YDLSKGQAEV
Alignment     :  :***   ::  :****:*:*   **    . **** *:.**:**.: ****** :*:

D. rerio      HSIKPWLFNI KAVAGQCPVI VVGTHADLCE ERHLQECLLK LQKELQSQPG
H. sapiens    DAMKPWLFNI KARASSSPVI LVGTHLDVSD EKQRKACMSK ITKELLNKRG
Alignment     .::******* ** *...*** :**** *:.: *:: : *: * : ***  .: *

D. rerio      FPAIRENHVL SACEESESLG RLRKAIYREL IGFKIQGQPV MGQLVPDCYV
H. sapiens    FPAIRDYHFV NATEESDALA KLRKTIINES LNFKIRDQLV VGQLIPDCYV
Alignment     *****: *.: .* ***::*. :***:* .*   :.***:.* * :***:*****

D. rerio      ELEKRLLQER SCAPADFPVL RHSRLMEILQ ETQLQLEEGE LPHAIHFLSE
H. sapiens    ELEKIILSER KNVPIEFPVI DRKRLLQLVR ENQLQLDENE LPHAVHFLNE
Alignment     **** :*.** .  .* :***:   :.**:::: *.****:*.* ****:***.*

D. rerio      AGVLLHFDDP VLQLKDLYFI DPQWFCRIIS QTLSLKSSGP WDSTKGVVQR
H. sapiens    SGVLLHFQDP ALQLSDLYFV EPKWLCKIMA QILTVKVEGC PKHPKGIISR
Alignment     :******:** .***.****: :*:*:*:*:: * *::* .*    . .**::.*

D. rerio      STVEKFVEKS RCFPKDHMIQ YFKLLEKFQI ALPFDHDQLL IPSSLSDHRP
H. sapiens    RDVEKFLSKK RKFPKNYMSQ YFKLLEKFQI ALPIGEEYLL VPSSLSDHRP
Alignment       ****:.*. * ***::* * ********** ***:..: ** :*********

D. rerio      VIELPHCENS EVIIRLYEMP YFPMGYWPRQ ISRLLEVSAF LLYGREKALK
H. sapiens    VIELPHCENS EIIIRLYEMP YFPMGFWSRL INRLLEISPY MLSGRERALR
Alignment     ********** *:******** *****:*.*  *.****:*.: :* ***:**:

D. rerio      PNRIYWRKGI YLSWSAEAYC LVEALTLEEN PASFIKITVP CSRKGRVLFG
H. sapiens    PNRMYWRQGI YLNWSPEAYC LVGSEVLDNH PESFLKITVP SCRKGCILLG
```

```
Alignment    ***:***:** **.**.**** ** : .*::: * **:***** ..*** :*:*

D. rerio     QVVDHIDSLL EEWFPGLLTT DIHGTGETLL KKWALYSFSD GQNCQKMLLE
H. sapiens   QVVDHIDSLM EEWFPGLLEI DICGEGETLL KKWALYSFND GEEHQKILLD
Alignment    *********: ********   ** * ***** ********.* *:: **:**:

D. rerio     DLLSNTNADG LLVNPEDPSC TLPISQISPD LVLSDQPSST ILDPEQLEME
H. sapiens   DLMKKAEEGD LLVNPDQPRL TIPISQIAPD LILADLPRNI MLNNDELEFE
Alignment    **:.::: .. *****::*   *:*****:** *:*:* * .  :*: ::**:*

D. rerio     LTAEYMLGDG GFGSVYKAVY KNEEVAVKIF NKHASALYVH RLVRQELAVL
H. sapiens   QAPEFLLGDG SFGSVYRAAY EGEEVAVKIF NKHTS----L RLLRQELVVL
Alignment     :.*::**** .*****:*.* :.******** ***:*      **:****.**

D. rerio     GRLCHPSLVG LLAAGCNPHI LVMELAPYGS LDSLFERENG SLSRKLQHRI
H. sapiens   CHLHHPSLIS LLAAGIRPRM LVMELASKGS LDRLLQQDKA SLTRTLQHRI
Alignment     :* ****:. ***** .*:: ******. ** ** *:::::. **:*.*****

D. rerio     ALHVADGLKY LHSSMIIYRD LKPHNVLLFN LKTDAEIVAK ITDYGIAQYC
H. sapiens   ALHVADGLRY LHSAMIIYRD LKPHNVLLFT LYPNAAIIAK IADYGIAQYC
Alignment    ********:* ***:****** *********. * .:* *:** *:********

D. rerio     CSMGVRSSEG TPGFRAPEVA RGNVIYNVQA DVYSFGLLLY DLLTYGERIS
H. sapiens   CRMGIKTSEG TPGFRAPEVA RGNVIYNQQA DVYSFGLLLY DILTTGGRIV
Alignment    * **:::*** ********** ******* ** ********** *:** * **

D. rerio     DGMKFPSEFD EVAVQGKLPD PVKDYGCSPW PEIESLMREC MRENPQDRPT
H. sapiens   EGLKFPNEFD ELEIQGKLPD PVKEYGCAPW PMVEKLIKQC LKENPQERPT
Alignment    :*:***.*** *: :****** ***:***:** * :*.*:::* ::****:***

D. rerio     SAQVFDRLNS AEMLCLTREL NVVGFPGECF VVSNSGGAAN GGKNPHVWIG
H. sapiens   SAQVFDILNS AELVCLTRRI LLP----KNV IVECMVATHH NSRNASIWLG
Alignment    ****** *** **::****.:  :     : . :*.   .: : ..:*. :*:*

D. rerio     GGSSSQKLGC VTAVDLETGG SLNQELDRSP ILCMVIIRAA DSCSDWLVAG
H. sapiens   CGHTDR--GQ LSFLDLNTEG YTSEEVADSR ILCLALVHLP VEKESWIVSG
Alignment     * :.:  *  :: :**:* *   .:*:  *  ***:.::: .  . ..*:*:*

D. rerio     SESGSLSIMD TINAKVLHRL KSVKDSVTSL YFHTELQHRC LKSYLLVGTA
H. sapiens   TQSGTLLVIN TEDGKKRHTL EKMTDSVTCL YCNSFSKQSK QKNFLLVGTA
Alignment    ::**:* ::: * :.*  * * :.:.****.* * ::  ::   *.:******

D. rerio     DGTLVIYEDS ALKLENGGPV KTLEVGDVNT PLMCLGPSSH PQERRSLWAA
H. sapiens   DGKLAIFEDK TVKLKGAAPL KILNIGNVST PLMCLSESTN STERNVMWGG
Alignment    **.*.*:**. ::**:...*: * *::*:*.* *****. *:: . **. :*..

D. rerio     CGTRIILFTV EFDVCRSIDT KPKPLFPLQA RVSGEACISR LAVDKHVYVS
H. sapiens   CGTKIFSFSN DFTIQKLIET RTSQLFSYAA FS--DSNIIT VVVDTALYIA
Alignment    ***:*: *:  :* : : *:* :.. **.  *     ::  *   :.**. :*::

D. rerio     KTGGHTVEVW DKKTERMVNL IDCMQLLGLS STRKPKVHSE DQSRPMVPSL
H. sapiens   KQNSPVVEVW DKKTEKLCGL IDCVHFLREV MVKE----NK ESKHKMSYSG
Alignment    * .. .**** *****:: .* ***:::*    .::    .: :..: *  *

D. rerio     VVKALLVQHS GTLWIGTRAG HILLVEVSSC HLLQTINPHC HSIRCMSSIL
H. sapiens   RVKTLCLQKN TALWIGTGGG HILLLDLSTR RLIRVIYNFC NSVRVMMTAQ
Alignment     **:* :*:.  :***** .* ****:::*: :*::.*  .* :*:* * :

D. rerio     LDTLNRKNVI LVLGRRQRIH LEQLKTQSGE DSVLTLWSSS LPLEARDLMR
H. sapiens   LGSL--KNVM LVLG-YNRKN TEGTQKQKEI QSCLTVWDIN LPHEVQNLEK
Alignment    *.:*  ***: ****  :* :  *  :.*.  :* **:*. . ** *.::* :
```

```
D. rerio      HCELRDKTTR RMRETLLN
H. sapiens    HIEVRKELAE KMRRTSVE
Alignment     * *:*.: :. :**.* ::
```

Figure 11. Multiple protein sequence alignment of Human and Zebrafish Lrrk2 orthologues. The alignment was performed by ClustalX2 ("*" identical residues, ":" conserved substitutions, "." semi-conserved substitutions). In yellow background is the Ankyrin domain in both sequences, the Leucine rich repeat domain in green background, ROC domain in Light blue. In grey background, the COR domain which is the domain with the highest homology between the two species. The Kinase domain is in dark blue and in red the WD40 domain.

The simple modular architecture research tool (SMART) analysis revealed that all the functional domains in Human Lrrk2 are also conserved in Zebrafish. The inter-species homology of the functional domains is different from domain to domain with the Ankyrin and WD40 ones having the lowest homology and the COR and Kinase domain the highest (Figure 12).

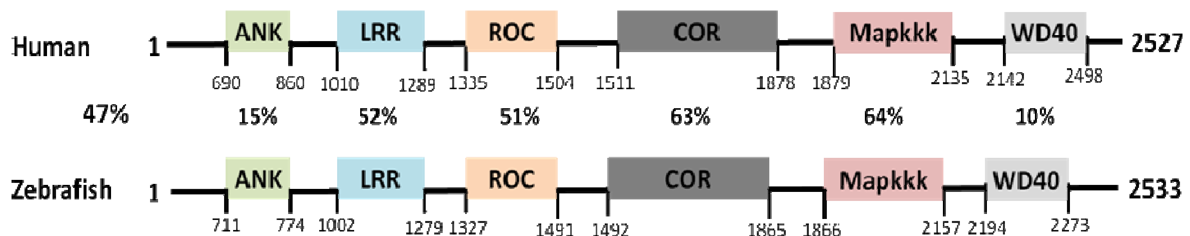


Figure 12. Human and Zebrafish Lrrk2 proteins. The two proteins have identical sizes and share the same functional domains with different homology between them. Homologies over 50% are found in four domains: LRR, ROC, COR and Kinase. The other two domains share very small homology with 15% for the Ankyrin domain and 10% for the WD40 domain.

4.2.3 Functional Domain of SNCAIP is 93% similar between species

Synphilin-1 (NP_005451.2) is the product of the gene *SNCAIP*. The blast analysis revealed one sequence in the Zebrafish genome with high homology, zgc:172291. The sequence from zebrafish is smaller (738 aa) when compared with the human one (938 aa) with 44% identity and 59% similiarity (Figure 13). The proteins share the same functional domain (Ankyrin) 78% identical between them and with a similarity of 93% (Figure 14). The Ankyrin repeats are responsible for protein-protein interactions and can be found in several types of proteins including cytoskeletal organizers and cell development and differentiation proteins (review at [91]).

```
D. rerio      MDVPEYLDLD EIDFTDDLPY SS---KSIPE LCRRHDGQND ERQALAINWP
H. sapiens    MEAPEYLDLD EIDFSDDISY SVTSLKTIPE LCRRCDTQNE DRSVSSSSWN
Alignment     *:.******* ****:**:.* *    *:*** **** * **: :*.. : .*

D. rerio      RSAASHSGAG LKPTGIADVY SKFRPVKRVS PLKHQPEETQ TQTETEGKSS
H. sapiens    CGISTLITNT QKPTGIADVY SKFRPVKRVS PLKHQPETLE NNESDDQKNQ
Alignment      . ::       ********* ********** *******  : .: . : *..

D. rerio      ---------E TDTPSGKEEP SKSRGLIN-- ------QALF GELEHYDLDM
H. sapiens    KVVEYQKGGE SDLGPQPQEL GPGDGVGGPP GKSSEPSTSL GELEHYDLDM
Alignment              * :*  .  :*  . . *: .        .: : **********

D. rerio      DEILDVPYIK CTQQAATLPR APS------- -------AGN THTLVHSESL
H. sapiens    DEILDVPYIK SSQQLASFTK VTSEKRILGL CTTINGLSGK ACSTGSSESS
Alignment     ********** .:** *::.: ..*               :*: : :   ***

D. rerio      SS-GTQFCVL SPV----MRK SKSADLRAQS LGFENTHEQI PDSKTG----
H. sapiens    SSNMAPFCVL SPVKSPHLRK ASAVIHDQHK LSTEETEISP PLVKCGSAYE
Alignment     **  : **** ***    :** :.:.    :. *. *:*. . *  * *

D. rerio      ---------- ---------- ---------- -------ADG GHASRKQARG
H. sapiens    PENQSKDFLN KTFSDPHGRK VEKTTPDCQL RAFHLQSSAA ESKPEEQVSG
Alignment                                           : .    ..:*. *

D. rerio      VSR------- EVDEEAKKSQ NILNIVRDGQ ISLLPHFAAE NLELIRDEDG
H. sapiens    LNRTSSQGPE ERSEYLKKVK SILNIVKEGQ ISLLPHLAAD NLDKIHDENG
Alignment     :.*        * .*  ** : .*****::** ******:**: **: *:**:*

D. rerio      NNLLHVSAAQ GHTDCLQHLT SLMGEDCLNE RNKQQLTPAG LSVRNGHLEC
H. sapiens    NNLLHIAASQ GHAECLQHLT SLMGEDCLNE RNTEKLTPAG LAIKNGQLEC
Alignment     *****::*:* **::****** ********** **.::***** *:::**:***

D. rerio      VRWMVSETEA IAELSCTREH PSLIHYAARY GQERVLLWLL QFMQEQAISL
H. sapiens    VRWMVSETEA IAELSCSKDF PSLIHYAGCY GQEKILLWLL QFMQEQGISL
Alignment     ********** ******:::. *******. * ***::***** ******.***

D. rerio      DEQDQNGNSA VHVAAQFGHL GCLQTLVEYG SNVTVQNQQC ERASQCAERQ
```

```
H. sapiens    DEVDQDGNSA VHVASQHGYL GCIQTLVEYG ANVTMQNHAG EKPSQSAERQ
Alignment     ** **:**** ****:*.*:* **:******* :***:**:   *:.**.****

D. rerio      GHTTCSRYLV VVETCMSLAS QVVKLTKQLH EQTTARVALQ NQLQLLLQTQ
H. sapiens    GHTLCSRYLV VVETCMSLAS QVVKLTKQLK EQTVERVTLQ NQLQQFLEAQ
Alignment     *** ****** ********** *********: ***. **:** **** :*::*

D. rerio      EPNGR--PPS PSCRVPPS-- -------DSW PEMTLTAEVA PENGQWVL--
H. sapiens    KSEGKSLPSS PSSPSSPASR KSQWKSPDAD DDSVAKSKPG VQEGIQVLGS
Alignment     :.:*:  *.* **.  .*:          *:  : . .:: .  ::*  **

D. rerio      ---------K QKHTETDGVM RKLLTKDAAE RTHPRDTHDA GAESGAGPGA
H. sapiens    LSASSRARPK AKDEDSDKIL RQLLGKEISE NVCTQEKLSL EFQDAQASSR
Alignment              *  *. ::* :: *:** *: :* .. .::. .    :.. ...

D. rerio      GPMKRLGVGE RRELKLARLK QIMQRSLSES DGDVYPPDE- ----TKHMSR
H. sapiens    NSKKIP--LE KRELKLARLR QLMQRSLSES DTDSNNSEDP KTTPVRKADR
Alignment     .. *     * :********: *:******** * *   .::      .:: .*

D. rerio      PTQLPIPEAE E--------- ---PKNTHSS SERKLSFTHR TSKSVDACNP
H. sapiens    PRPQPIVESV ESMDSAESLH LMIKKHTLAS GGRRFPFSIK ASKSLDGHSP
Alignment     *   ** *:  *              *:* :* . *::.*: : :***:*. .*

D. rerio      SPSSDQSDPE TR-------- ---------- ------TEAG DKVTTSPKSA
H. sapiens    SPTSESSEPD LESQYPGSGS IPPNQPSGDP QQPSPDSTAA QKVATSPKSA
Alignment     **:*:.*:*:  .                           : *. :**:******

D. rerio      LKSPSSRRKT SQNLKLRVTF DEPP------ ------RKDG AAGDTKAPST
H. sapiens    LKSPSSKRRT SQNLKLRVTF EEPVVQMEQP SLELNGEKDK DKGRTLQRTS
Alignment     ******:*:* ********** :**             .**    * *   ::

D. rerio      KE-------K RPFGAFRSIM ETLSGNQNNN NSNSSSSNAQ -------SPG
H. sapiens    TSNESGDQLK RPFGAFRSIM ETLSGNQNNN NNYQAANQLK TSTLPLTSLG
Alignment     ..       * ********** ********** *. .::.: :        * *

D. rerio      KHTGKKSKSK TSAV------ -
H. sapiens    RKTDAKGNPA SSASKGKNKA A
Alignment     ::*. *.:.  :**
```

Figure 13. Alignment of the Human and Zebrafish Synphilin-1 sequences. The alignment was performed by ClustalX2 ("*" identical residues, ":" conserved substitutions, "." semi-conserved substitutions). In yellow background is the only functional domain of Synphilin-1: Ankyrin.

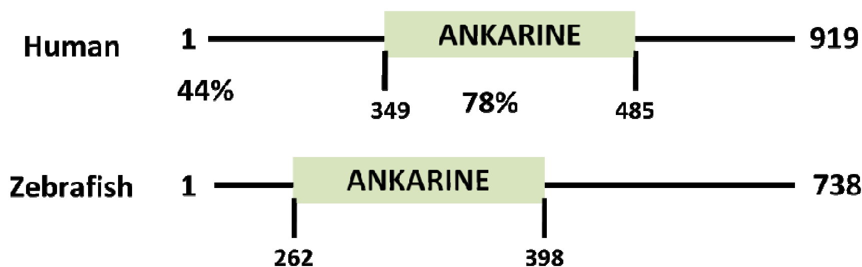


Figure 14. Human and Zebrafish synphilin-1. Although the total protein sequence is only 44% homologue between species, the functional domain shares 78% identity and 93% similarity.

4.2.4 Glucocerebrosidase in zebrafish has the same functional domain found in humans

The blast of Glucocerebrosidase sequence (NP_001005750.1) against the fish model genome showed a highly homologous sequence in zebrafish, LOC559072. Both sequences share 54% identity, 68% similarity, and similar sizes, 536 in Human GBA and 518 in Zebrafish (Figure 15). The functional domain that represents the majority of the sequence is 57% identical (Figure 16).

```
D. rerio      ---------- -MRETALFIL LAGIIT---- ---TARADSK DGCLALNFGH
H. sapiens    MEFSSPSREE CPKPLSRVSI MAGSLTGLLL LQAVSWASGA RPCIPKSFGY
Aligment                   :  : . : :** :*        .: *..    *:. .**:

D. rerio      GSVVCVCNAT YCDSLGRTVL PDAGQFLSYV SNKAGSRLME SQGQFQKNST
H. sapiens    SSVVCVCNAT YCDSFDPPTF PALGTFSRYE STRSGRRMEL SMGPIQANHT
Aligment      .********* ****:. ..: *  * *  *  *.::* *:   * * :* * *

D. rerio      GAALRITLNP SQKFQHIKGF GGAMTDAAAI NILSLSSGAQ DQLLRQYFST
H. sapiens    GTGLLLTLQP EQKFQKVKGF GGAMTDAAAL NILALSPPAQ NLLLKSYFSE
Aligment      *:.* :**:* .****::*** *********: ***:**. ** : **:.***

D. rerio      DGIEYRFVRV PVASCDFSTR LYTYADTPED YDLQNFTLAK EDVHMKIPLL
H. sapiens    EGIGYNIIRV PMASCDFSIR TYTYADTPDD FQLHNFSLPE EDTKLKIPLI
Aligment      :** *.::** *:****** *  *******:* ::*:**:*.: **.::****:

D. rerio      QRAQALSAQP LYLFASAWSA PAWLKTNGAL IGKGSLKGKP GGKEHKTWAQ
H. sapiens    HRALQLAQRP VSLLASPWTS PTWLKTNGAV NGKGSLKGQP GDIYHQTWAR
Aligment      :**  *: :* : *:**.*:: *:*******:  *******:* *.  *:***:

D. rerio      YYIRFLEEYR KYNLSFWGLT SGNEPTAGEM TNYSFQALGF TPETQRDWIA
H. sapiens    YFVKFLDAYA EHKLQFWAVT AENEPSAGLL SGYPFQCLGF TPEHQRDFIA
Aligment      *:::**: *  :::*.**.:* : ***:** : :.*.**.*** *** ***:**

D. rerio      LDLGPALHSS SFSKTQLMIL DDNRLMLPHW AKVVLSDIKA ARYVHGIGVH
H. sapiens    RDLGPTLANS THHNVRLLML DDQRLLLPHW AKVVLTDPEA AKYVHGIAVH
Aligment       ****:* .* :. :.:*::* **:**:**** *****:* :* *:*****.**

D. rerio      WYFDRLVPPD VTLTSTHHLY PDYFLFATEA CAGWSPVDRG VRLGSWDRAE
H. sapiens    WYLDFLAPAK ATLGETHRLF PNTMLFASEA CVGSKFWEQS VRLGSWDRGM
Aligment      **:* *.*.. .** .**:*: *: :***:** *.* .  ::. ********.

D. rerio      DYAHDIIQDL NNYVTGWTDW NLALNQDGGP NWVKNFVDSP IIVDPSKDIF
H. sapiens    QYSHSIITNL LYHVVGWTDW NLALNPEGGP NWVRNFVDSP IIVDITKDTF
Aligment      :*:*.** :*   :*.***** ***** :*** ***:****** **** :** *

D. rerio      YKQPTFYSMA HFSKFLWEES QRVGVSFSQQ TSLEMSAFIR PDASAVLIIL
H. sapiens    YKQPMFYHLG HFSKFIPEGS QRVGLVASQK NDLDAVALMH PDGSAVVVVL
Aligment      **** ** :. *****: * * ****:  **: ..*:  *::: **.***:::*
```

```
D. rerio      NRSEEEVPFE VWDQTVGFLP GSAPPHSILT LLWNRQ
H. sapiens    NRSSKDVPLT IKDPAVGFLE TISPGYSIHT YLWRRQ
Aligment      ***.::**:   :  *  :****     :*  :** *   **.**
```

Figure 15. Multiple protein sequence alignment of Human GBA and its orthologue in Zebrafish. The alignment was performed by ClustalX2 ("*" identical residues, ":" conserved substitutions, "." semi-conserved substitutions). In yellow background is the Glyco domain that is common in both sequences.

.

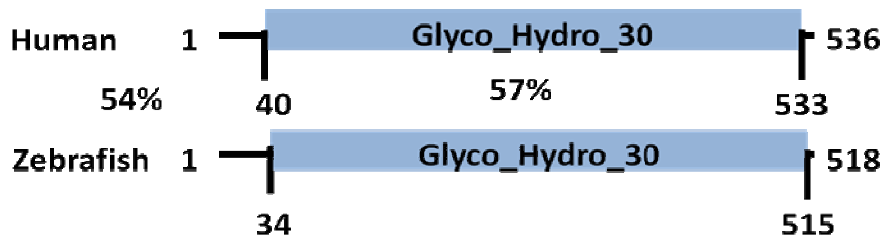


Figure 16. Human and Zebrafish GBA protein sequence homology and functional domain. The proteins share the same size and are 54% identical. The functional domain represents the majority of the protein sequence and, therefore, the homology is very similar to the total protein sequence, 57%.

4.3 ATP13a2 is expressed during Zebrafish embryonic development

Using the WISH technique, a spatial analysis of ATP13a2 was performed. Embryos at different stages of development were probed: 12, 24, 48, 72 hpf and 6 dpf. It is possible to see that the gene expression is more ubiquitous in the early development stages, 12hpf, while with ageing it localizes specifically in the embryo's brain (Figure 17 A – E). The sense probe, which works as a control probe, didn't provide any signal (Figure 17: F and G).

Figure 17. ATP13a2 expression during embryonic development. Gene expression is marked in purple as a result of a reaction with the antibody of digoxigenine. **A)** at 12hpf, ATP13a2 is found throughout the embryo**; B)** The expression, at 24hpf, is only found in the upper part of the embryo, especially in the brain area; **C)** Although the expression is more diffuse, a strong signal in the brain at 48 hpf is still observed. The expression is very similar at 72 hpf **D)** and 6 dpf **E)** being restricted to the brain of the embryo, more specifically at the front end. **F) and G)** No expression is observed with the sense probe in embryos with 12 hpf and 48 hpf, for example. Scale bar: 100µm.

5. Discussion

5.1 **Transgenic zebrafish as a model of PD**

Three tg lines were successfully generated: THrat:IRES:GFP, THrat:a-syn:IRES:GFP and THzeb:GFP. Currently, these lines are in different stages of development with the TH:IRES:GFP and TH:a-syn:IRES:GFP in process of development. The fish injected with the construct THzeb:GFP were already crossed with wt zebrafish and positive F1 generation was obtained. Two of them, THrat:IRES:GFP and THzeb:GFP, will be used as control lines: the first will be a control for the THrat:a-syn:IRES:GFP, specially for behavior analysis, to ensure that any result is only due to the expression of a-syn and not from the micro-injection process.

Other model's goal is to study the advantage/disadvantage of using the IRES sequence. The THzeb:GFP will be compared with THzeb:IRES:GFP, which is under development, thus it will be possible to understand the effect of IRES. If IRES is proven to negatively modulate the expression of the gene throughout the fish development, it will be necessary to find new alternatives. The simplest one would be the fusion of a-syn with GFP, however it is controversial if the normal role of a-syn is maintained and, therefore, other alternatives must be found.

Different concentrations of the co-injection tranposase mRNA and DNA concentrations were investigated with the THrat:IRES:GFP construct. With initial concentrations of 30ng/µl DNA: 25ng/µl mRNA transposase no strong fluorescence was detected (Figure 6 A-B) when compared to the control (non-injected) Zebrafish (Figure6 I-J). Higher concentrations were tested and the best one used to generate the transgenic lines was 80ng/µl DNA: 50 ng/µl mRNA transposase. With this concentration it is possible to see visible expression of the construct as seen in Figure 6 C-D). The pattern of expression, being mostly present in the zebrafish eye, is in accordance with the expression seen by Gao *et al* [92].This concentration is the maximum possible concentration since higher ones have induced morphological abnormalities in the fish.

Concerning the THzeb:GFP construct, lower concentrations were sufficient to generate the founder fish (Figure 6 G-H).

To understand the pattern and levels of a-syn expression and compare them with the fluorescence analysis results, behavior tests and molecular analysis by western blot and immunohistochemistry will be investigated in the F1 generation.

So far, no molecular tests will be performed in the founder fish for two main reasons. The first is the possibility to obtain false positives. Western-blot analysis would indicate the a-syn levels, however the construct might only be transiently expressed and not inserted in the genome. If a-syn is found in the next generations, it is ensured that our gene of study is integrated in the zebrafish genome and therefore a transgenic line has been generated. The second reason relates to the number of founder fish. The efficiency of the Tol2 technique is between 30% to 70% and from those remaining embryos only the ones with the higher fluorescence were chosen to become founder fish for the lines.

5.2 **PD genes in Zebrafish**

In agreement with previous studies [68-72], the studied genes revealed high homology with the human orthologues. For ATP13a2-5, Zebrafish has only one matching sequence. Through phylogenetic and functional domains analysis, it is possible to determine that the hit, zgc:136762 is more identical to ATP13a2 than the other ATP13ases human family members and therefore might be considered as zebrafish ATP13a2. Both functional domains are conserved meaning that the function of the protein might be conserved in the two species. The WISH experiments suggested that in the early stages of development, 12 hpf, ATP13a2 is present throughout the embryo's body (Figure 17.A), probably due to maternal heritance rather than embryonic expression. The expression pattern during development is more restrained in the upper part of the body after 24 hpf (Figure 17.B), and embryos older than 48 hpf show a specific expression in the brain, especially in the frontal area (Figure 17.C-E).

Lrrk2 is one of the most important protein linked to PD. Due to its size, it has been one of the most difficult proteins to study in animal models. The sequence matching to

human Lrrk2 is a 2533 amino-acid protein sharing the same functional domains. The homology between the different domains is very variable ranging from 10 in the WD40 domain to 64% in the kinase domain. In agreement with recent results, the deletion of the WD40 domain was also shown to trigger a Parkinsonian-like behavior in Zebrafish [93].

Zebrafish has also a Synphilin-1 homologue and is in this protein where the homology between the two species is more remarkable. Although the homology of the total sequence is 44%, the functional domain is conserved between the two species with 78% identity and 93% of similarity. Synphilin-1 is known to interact with a-syn and might contribute to the formation of protein aggregates [94], suggesting that, although a-syn has not yet been described in zebrafish, it may be present in the zebrafish genome similarly to other fish species [95, 96]. The continuous work to sequence the full genome of Zebrafish will, probably, allow the discovery of a-syn gene. It would also be interesting to analyze the role of synphilin-1 in zebrafish and develop an in vivo model to study protein aggregation, commonly reported in PD patients.

Glucocerebrosidase is also present in the zebrafish genome. The percentages of homology are in agreement with the average of values presented by the other studied proteins, with 54% and the similarity of 68%. Both human and zebrafish have the same functional domain, Glyco_Hydro_30, which is responsible for the regulation of beta-glucocerebrosidase. Moreover, the values of identity and similarity are identical to the whole sequence values, 57% and 68% respectively.

The high homology of all PD-related proteins between Zebrafish and Human (Table 2) a homology that is present also in their conserved domains (Table 3) makes zebrafish a good animal to study the disease.

Table 2. Identity and Similarity between human PD genes and Zebrafish		
Gene	**Identity**	**Similarity**
Parkin	62%	75%
Pink – 1	54%	(67%)*
DJ-1	83%	89%
UCHL-1	(66%)*	79%
ATP13a2	50%	69%
Lrrk2	41%	57%
Synphilin – 1	44%	59%
Glucocerebrosidase	54%	68%
a-syn	Nothing	Nothing

Table 3- Identity and similarity of the studied genes Functional Domains			
Gene	**Functional Domain**	**Identity**	**Similarity**
ATP13a2	E1_E2 ATPase	53%	66%
	HAD_like	51%	69%
LRR2K2	Ankyrin	15%	24%
	LRR	52%	72%
	ROC	51%	76%
	COR	63%	79%
	Mapkkk	64%	75%
	WD40	10%	15%
Synphilin -1	Ankyrin	78%	93%
Glucocerebrosidase	Glyco_Hydro	57%	68%

6. Conclusion

This project, using transgenic zebrafish lines, opens new doors for the study of PD.
The purpose of this thesis was to demonstrate the viability of Zebrafish as a model to study PD and to develop several transgenic lines to further study the hallmarks of PD in this organism.
Three different tg lines have been successfully initiated. As a starting point, we have created two control lines: THzeb:GFP and THrat:IRES:GFP. These control lines serve different purposes. They will be used as controls of other developed lines with the constructs. THzeb:IRES:GFP, to analyze the role of IRES sequence, and THrat:a-syn:IRES:GFP to study the effect of a-syn, respectively. The later construct, THrat:a-syn:IRES:GFP, has already been injected in zebrafish embryos to further study the role of a-syn.
Bioinformatic analysis of PD associated genes showed that the homology of human and Zebrafish orthologues is greater than 40%, and the functional domains are conserved between them in all studied proteins. *Danio rerio* has only one gene orthologue to the human ATP13a2-ATP13a5 proteins, phylogenetically closer to ATP13a2. The results from our WISH studies demonstrated that this gene is expressed during embryogenesis. Regarding Lrrk2, there is a high range of identity between the different domains between 10 to 63%, with the more conserved domains being responsible for the enzymatic activity of this protein. From the studied proteins, Synphilin-1, with its 93% similarity in the functional domain, is the more conserved protein during evolution. For GBA, the homology of both total sequence and functional domain is around 50%.
Altogether, these results suggests that zebrafish can be a good animal model to study PD, and a remarkable *in vivo* tool to better understand the molecular mechanisms and behavior patterns underlying PD.

7. Future Work

Two new constructs are being developed: THzeb:IRES:GFP and THzeb:a-syn:IRES:GFP. These two lines are comparable between them, and we will analyze the effect of a-syn under the control of a promoter specific for zebrafish. The comparison of the first construct with the THzeb:GFP, already generated, will enable to investigate if the IRES sequence influences the construct expression. If so, new alternatives for the generation of the constructs must be studied.

In addition, two other constructs will be generated with the Nestin promoter. This promoter, contrary to tyrosine hydroxylase that is specific for dopaminergic neurons, is ubiquitously expressed in all the Zebrafish brain. This will provide important knowledge on the effect of the human a-syn expression in all the brain. These constructs will be developed after the analysis of the effect of the IRES sequence.

For the F1 generation, biochemical and molecular studies will be performed. More specifically, the expression of a-syn will be analyzed by western-blot and by immunofluorescence. From this generation on, behavioral tests will be performed to observe both the effects of a-syn overexpression and also the effects of specific parkinsonism-inducible drugs like MPTP in these lines.

Concerning the bioinformatics studies, new WISH for Lrrk2 will be performed with different probes. Probes for Synphilin-1 and GBA are being developed to perform WISH in the same stages of embryonic development.

As a final goal, we will probe for Zebrafish a-syn using degenerate primers, and study its expression pattern during embryonic development to adult fish. The presence of a 12 amino-acid constant sequence specific for a-syn between several species facilitates the use of this technique.

8. Acknowledgments

Thanks to:

- My supervisors Dr. Tiago Fleming Outeiro and Dr. Ana Dulce Correia for the opportunity and the support throughout this Thesis.
- The UNCM team for all the help, support and good times. Special thanks to Sandra Tenreiro, Hugo Miranda and Susana Gonçalves.
- Dr. Ana Sebastião for all the help and practical solutions along this two years of Master course
- The UDEV group and specially Raquel Lourenço for all the help and patience in the in-situ experiments.
- Nuno Afonso from TechnoZeb in the Bioinformatics and in the in-situ experiments.
- Lara Carvalho for all the help in the Fish Facility
- Last but not least, my family, friends and Rita for supporting me in all stages of my life and without them this wouldn't be possible.

9. References

1. EC, H., *Biochemistry of Parkinson's disease with special reference to the dopaminergic system.* Molecular Neurobiology 1994. **9**: p. 135-42
2. WP, G., *Age-related loss of dorsal vagal neurons in Parkinson's disease.* Neurology, 1992. **42**: p. 2106-11.
3. Carr, J., *Familial and sporadic Parkinson's disease usually display the same clinical features.* Parkinsonism & Related Disorders 2003. **9**: p. 201-4.
4. Papapetropoulos, S., *A Prospective Study of Familial versus Sporadic Parkinson's Disease.* Neurodegenerative diseases, 2007. **4**: p. 424-7
5. Chandra, S., *Alpha-synuclein cooperates with CSPalpha in preventing neurodegeneration.* Cell, 2005. **4**: p. 383-96.
6. Outeiro, T., *Formation of Toxic Oligomeric α-Synuclein Species in Living Cells.* PLoS ONE 2008. **3**.
7. Rajagopalan, S., *Alpha synuclein aggregation: is it the toxic gain of function responsible for neurodegeneration in Parkinson's disease?* Mechanism of ageing and development, 2001. **122**: p. 1499-510.
8. Polymeropoulos, M., et al., *Mutation in the alpha-synuclein gene identified in families with Parkinson's disease.* Science, 1997. **276**: p. 2045–2047.
9. Stefanis, L., et al., *Expression of A53T Mutant But Not Wild-Type alpha-Synuclein in PC12 Cells Induces Alterations of the Ubiquitin-Dependent Degradation System, Loss of Dopamine Release, and Autophagic Cell Death.* . The Journal of Neuroscience, 2007. **21**: p. 2045–47
10. Jensen, P., et al., *Binding of alpha-synuclein to brain vesicles is abolished by familial Parkinson's disease mutation.* The Journal of Biological Chemistry, 1998. **273**: p. 26292–94.
11. Outeiro, T.F. and S. Lindquist, *Yeast Cells Provide Insight into Alpha-Synuclein Biology and Pathobiology.* Science, 2003(302): p. 1772-1775.
12. Song, W., et al., *The Parkinson disease-associated A30P mutation stabilizes alpha-synuclein against proteasomal degradation triggered by heme oxygenase-1 over-expression in human neuroblastoma cells.* Journal of Neurochemistry, 2009. **110**: p. 719-33.
13. Greenbaum, r.A., et al., *The E46K Mutation in α-Synuclein Increases Amyloid Fibril Formation.* The Journal of Biological Chemistry, 2005. **280**: p. 7800 – 7.
14. Zarranz, J., et al., *The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia.* Annals of Neurology, 2004. **55**: p. 153-6.
15. Singleton, A., et al., *alpha-Synuclein locus triplication causes Parkinson's disease.* Science, 2003. **31**.
16. Lavedan, C., *The synuclein family.* Genome Research, 1998. **8**: p. 871-80
17. Chandra, S., et al., *A broken α-helix in folded α-synuclein. Journal of Biological Chemistry* 2003. **278**: p. 15313–18
18. Cai, H., K. Reinisch, and S. Ferro-Novick, *Coats, Tethers, Rabs, and SNAREs Work Together to Mediate the Intracellular Destination of a Transport Vesicle.* Developmental Cell, 2007. **12**: p. 671-82.
19. Cabin, D.E., et al., *Synaptic vesicle depletion correlates with attenuated synaptic responses to prolonged repetitive stimulation in mice lacking α-synuclein.* Journal of Neuroscience, 2002. **22**: p. 8797-807

20. Murphy, D.D., et al., *Synucleins are developmentally expressed, and α-synuclein regulates the size of the pre- synaptic vesicular pool in primary hippocampal neurons.* Journal of Neuroscience, 2000. **20**: p. 3214-20
21. Larsen, K.E., et al., *α-Synuclein Overexpression in PC12 and Chromaffin Cells Impairs Catecholamine Release by Interfering with a Late Step in Exocytosis.* Journal of Neuroscience, 2006. **26**: p. 11915-922.
22. Davidson, W.S., et al., *Stabilization of alpha-synuclein secondary structure upon binding to synthetic membranes.* Journal of Biological Chemistry, 1998. **273**: p. 9443–9.
23. Gedalya, T.B., et al., *α-Synuclein and Polyunsaturated Fatty Acids Promote Clathrin-Mediated Endocytosis and Synaptic Vesicle Recycling* Traffic, 2009. **10**: p. 218-34.
24. Cookson, R.M., *α-Synuclein and neuronal cell death.* Molecular Neurodegeneration, 2009. **4**.
25. Epand, R.M., *The Amphipathic Helix*. 1993, CRC Press
26. Funayama, M., H. K, and K. H, *A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1.* Annals of Neurology, 2002. **51**.
27. Stenmark, H. and V. Olkkonen, *The Rab GTPase family.* Genome Biology 2001. **301**: p. 1077-87.
28. Gloeckner, C., et al., *The Parkinson disease-associated protein kinase LRRK2 exhibits MAPKKK activity and phosphorylates MKK3/6 and MKK4/7, in vitro.* Journal of Neurochemistry, 2009. **109**.
29. Mata, I., K. JM, and T. JP, *Lrrk2 pathogenic substitutions in Parkinson's disease.* Neurogenetics, 2005. **6**: p. 171–77.
30. Healy, D., F. M, and O.S. SS, *Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case–control study.* The Lancet Neurology, 2008. **7**: p. 583–90
31. Zimprich, A., B. S, and L. P, *Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology.* Neuron, 2004. **44**: p. 601-7.
32. Biskup, S., M. DJ, and R. A, *Dynamic and redundant regulation of LRRK2 and LRRK1 expression.* BMC Neuroscience, 2007. **8**.
33. White, L., et al., *MAPK-pathway activity, Lrrk2 G2019S, and Parkinson's disease.* Journal of Neuroscience Research, 2007. **85**: p. 1288–94.
34. Li, X., et al., *Leucine-rich repeat kinase 2 (LRRK2)/PARK8 possesses GTPase activity that is altered in familial Parkinson's disease R1441C/G mutants.* Journal of Neurochemistry, 2007. **103**: p. 238–47.
35. Westerlund, M., B. AC, and A. A, *Developmental regulation of leucine-rich repeat kinase 1 and 2 expression in the brain and other rodent and human organs: implications for Parkinson's disease.* Neuroscience, 2008. **152**: p. 429–36.
36. Higashi, S., et al., *Expression and localization of Parkinson's disease-associated leucine-rich repeat kinase 2 in the mouse brain.* Journal Neurochemestry, 2006. **100**: p. 368–81.
37. Greggio, E., J. S, and K. A, *Kinase activity is required for the toxic effects of mutant LRRK2/dardarin.* Neurobiology of Disease, 2006. **23**: p. 329–41
38. Gloeckner, C., K. N, and S. A, *The Parkinson disease causing LRRK2 mutation I2020Tis associated with increased kinase activity.* Human Molecular Genetics 2006. **15**: p. 223–32.
39. West, A., M. DJ, and B. S, *Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity.* Proceedings of the National Academy of Sciences, 2005. **102**: p. 16842–47
40. Hatano, T., K. S, and I. S, *-rich repeat kinase 2 associates with lipid rafts.* Human Molecular Genetics, 2007. **16**: p. 678–90
41. Smith, W., P. Z, and J. H, *Kinase activity of mutant LRRK2 mediates neuronal toxicity.* Nature Neuroscience, 2006. **9**: p. 1231–33

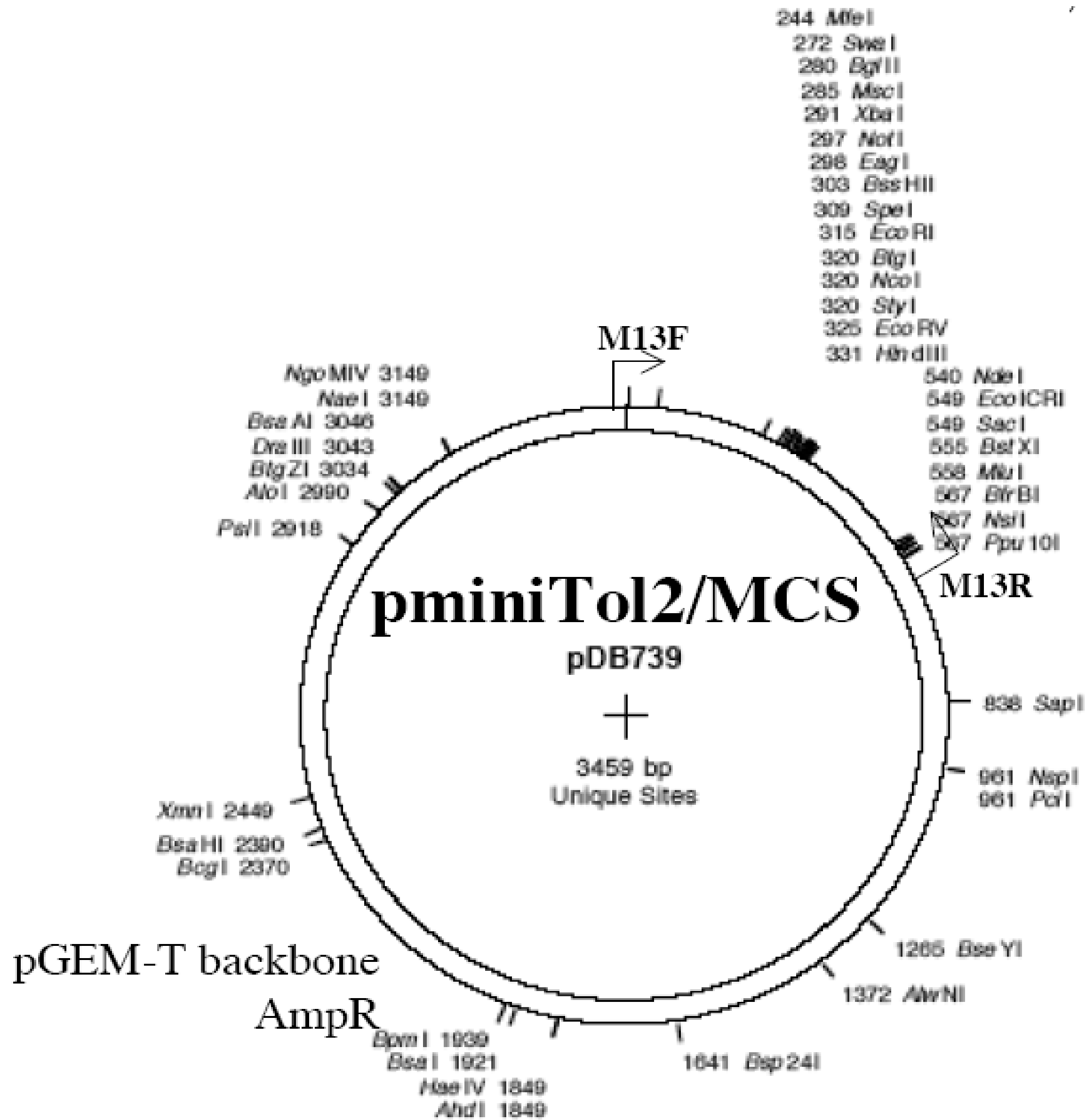
42. Ito, G., T. Okai, and G. Fujino, *GTP binding is essential to the protein kinase activity of LRRK2, a causative gene product for familial Parkinson's disease.* Biochemistry, 2007. **46**: p. 1380–88.
43. Schultheis, P.J., et al., *Characterization of the P5 subfamily of P-type transport ATPases in mice.* Biochemical and Biophysical Research Communications, 2004. **323**: p. 731-38.
44. Moller, A.B., et al., *Phylogenetic analysis of P5 P-type ATPases, a eukaryotic lineage of secretory pathway pumps.* Molecular Phylogenetics and Evolution, 2008. **46**: p. 619-34.
45. Ramirez, A., et al., *Hereditary parkinsonism with dementia is caused by mutations inATP13A2, encoding a lysosomal type 5 P-type ATPase.* Nature Genetics, 2006. **38**: p. 1184-91.
46. Wakabayashi, K., et al., *Synphilin-1 is present in Lewy bodies in Parkinson.* Annals of Neurology, 2000. **47**: p. 521 – 3.
47. Kawamata, H., et al., *Interaction of alpha-synuclein and synphilin-1: effect of Parkinson's disease-associated mutations.* Journal Neurochemestry, 2001. **77**: p. 929 – 34.
48. Engelender, S., et al., *Synphilin-1 associates with alpha-synuclein and promotes the formation of cytosolic inclusions.* Nature Genetics, 1999. **22**: p. 110-4.
49. Chung, K.K., et al., *Parkin ubiquitinates the alpha-synuclein-interacting protein, synphilin-1: implications for, Lewy-body formation in Parkinson's disease.* Nature Medicine, 2000. **7**: p. 1144-50.
50. Smith, W.W., et al., *LRRK2 interacts with Synphilin-1.* 2006.
51. Marx, F.P., et al., *Identification and functional characterization of a novel R621C mutation in the synphilin-1 gene in Parkinson.* Human Molecular Genetics, 2003. **12**: p. 1223-31.
52. Tayebi, N., et al., *Gaucher disease with parkinsonian manifestations: does glucocerebrosidase deficiency contribute to a vulnerability to parkinsonism?* Molecular Genetics and Metabolism, 2003. **79**: p. 104–9.
53. Machaczka, M., et al., *American Journal of Hematology.* Parkinson syndrome preceding clinical manifestation of Gaucher disease, 1999. **6**: p. 216-7.
54. Lwin, A., et al., *Glucocerebrosidase mutations in subjects with parkinsonism.* Molecular Genetics and Metabolism, 2004. **81**: p. 70-3.
55. Driscoll, M. and B. Gerstbrein, *Dying for a cause: invertebrate genetics takes on human neurodegeneration. .* Nature Reviews Genetics, 2003. **4**: p. 181-94.
56. Willingham, S., et al., *Yeast genes that enhance the toxicity of a mutant huntingtin fragment or alpha-synuclein.* Science, 2003. **302**: p. 1769-72.
57. Auluck, P., et al., *Chaperone suppression of alpha-synuclein toxicity in Drosophila model for Parkinson´s disease.* Science, 2002. **295**: p. 865-68
58. RECCHIA, A., et al., *alpha-Synuclein and Parkinson's disease.* Faseb Journal, 2004. **18**: p. 617-26.
59. Melrose, H.L., et al., *Parkinson's disease: a rethink of rodent models.* Experimental Brain Research, 2006. **173**: p. 196-204.
60. Kirik, D., et al., *Parkinson-like neurodegeneration induced by targeted overexpression of alpha-synuclein in the nigrostriatal system.* Journal Neuroscience, 2002. **22**: p. 2780-91.
61. Schneider, B., Z. R., and P. Aebischer, *Viral vectors, animal models and new therapeis for Parkinson´s disease.* Parkinsonism & Related Disorders, 2008. **14**: p. 169-71.
62. Wulliman, M.F., B. Rupp, and H. Reichert, *Neuroanatomy of the Zebrafish Brain: A Topological Atlas*. 1996: Birkhäuser Basel.
63. Lieschke, G.J. and P.D. Currie, *Animal models of human disease: zebrafish swim into view.* Nature Reviews Genetics, 2007. **8**: p. 353-367.

64. Bai, Q., et al., *Generation of the transgenic zebrafish model of Tauopathy using a novel promoter element derived fro the zebrafish eno2 gene. .* Nucleic acid research 2007. **19**: p. 6501-6516
65. Ma, P., *Catecholaminergic systems in the zebrafish. IV.Organization and projection pattern of dopaminergic neurons in the diencephalon.* The Journal of Comparative Neurology, 2003. **460**: p. 13–37.
66. Koulen, P., et al., *Conservation of localization patterns of IP(3) receptor type 1 in cerebellar Purkinje cells across vertebrate species. .* Journal Neuroscience Research 2000. **61**: p. 493–99
67. Westerfield, M., J.V. McMurray, and J.S. Eisen, *Identified motoneurons and their innervation of axial muscles in the zebrafish.* Journal Neuroscience 1986. **6**: p. 2267–77
68. Bai, Q., et al., *Zebrafish DJ-1 is evolutionarily conserved and expressed in dopaminergic neurons.* Brain Research 2006. **1113**: p. 33-44
69. Flinn, L., et al., *Complex I deficiency and dopaminergic neuronal cell loss in parkin-deficient zebrafish.* Brain Research, 2009. **132**: p. 1613-23
70. Anichtchik, O., et al., *Loss of PINK1 Function Affects Development and Results in Neurodegeneration in Zebrafish. .* Neurobiology of Disease, 2008. **28**: p. 8199-8207.
71. Son, O., et al., *Cloning and expression analysis of a Parkinson's disease gene, UCH-L1, and its promoter in zebrafish.* Biochemical and Biophysical Research Communications, 2003. **312**: p. 601-07.
72. Sun, Z. and A. Gitler, *Discovery and characterization of the three novel synucelin genes in zebrafish.* Developmental Dynamics, 2008. **237**: p. 2490-95
73. Holzschuh, J., et al., *Dopamine transporter expression distinguishes dopaminergic neurons from other catecholaminergic neurons in the developing zebrafish embryo.* Mechanisms of Development 2001. **101**: p. 237-43.
74. Panula, P., et al., *Modulatory Neurotransmitter Systems and Behavior: Towards Zebrafish Models of Neurodegenerative Diseases.* Zebrafish, 2006. **3**: p. 235-47
75. Kawakami, K. and A. Shima, *Identification of the Tol2transposase of the medaka fish Oryzias latipesthat catalyzes excision of anonautonomous Tol2element in zebrafish Danio rerio. .* Gene, 1999. **240**: p. 239-44
76. Kawakami, K., A. Shima, and N. Kawakami, *Identification of a functional transposase of the Tol2element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage.* Proceedings of the National Academy of Sciences, 2000. **97**: p. 11403-08.
77. Urasaki, A., G. Morvan, and K. Kawakami, *Functional dissection of the Tol2transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition.* Genetics, 2006. **174**: p. 639-49.
78. Kawakami, K., et al., *A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. .* Developmental Cell, 2004. **7**: p. 133-44
79. Kawakami, K., *Tol2: a versatile gene transfer vector in vertebrates.* Genome Biology, 2007. **8**.
80. Lin, S., et al., *Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish.* Science, 1994. **265**: p. 666-9.
81. Gaiano, N., et al., *Highly efficient germ-line transmission of proviral insertions in zebrafish.* Proceedings of the National Academy of Sciences, 1996. **93**: p. 7777-82
82. Hubbard, T.J.P., et al., *Ensembl 2009.* Nucleic Acids Research, 2009. **37**.
83. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.
84. Hall, T.A., *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.* Nucleic Acids Symposium Series, 1999. **41**: p. 95-8.
85. Letunic, I., T. Doerks, and P. Bork, *SMART 6: recent updates and new developments.* Nucleic acid research, 2008. **2008**.

86. Marchler-Bauer, A., et al., *CDD: specific functional annotation with the Conserved Domain Database.* Nucleic acid research, 2009.
87. Rozen, S. and H.J. Skaletsky, *Primer3 on the WWW for general users and for biologist programmers.*, in *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, M.S.e. Krawetz S, Editor. 2000, Humana Press: Totowa, NJ,.
88. Thisse, C. and B. Thisse, *High-resolution in situ hybridization to whole-mount zebrafish embryos. .* Nature Protocols, 2008. **3**: p. 59-69
89. Axelsen, K.B. and M.G. Palmgren, *Evolution of Substrate Specificities in the P-Type ATPase Superfamily.* Journal of Molecular Evolution, 1998(46): p. 84-101.
90. Ugochukwu, E., et al., *The crystal structure of human Haloacid Dehalogenase-like Hydrolase Domain containing 1A (HDHD1A)*. 2010.
91. Li, J., A. Mahajan, and M.-D. Tsai, *Ankyrin Repeat: A Unique Motif Mediating Protein-Protein Interactions.* Biochemistry, 2006. **45**: p. 15168-78.
92. Gao, Y., P. Li, and L. Li, *Transgenic Zebrafish That Express Tyrosine Hydroxylase Promoter in Inner Retinal Cells.* Developmental Dynamics, 2005. **233**: p. 921-29.
93. Sheng, D., et al., *Deletion of the WD40 Domain of LRRK2 in Zebrafish Causes Parkinsonism-Like Loss of Neurons and Locomotive Defect* PLoS Genetics, 2010. **6**: p. e1000914.
94. Kawamata, H., et al., *Interaction of α-synuclein and synphilin-1: effect of Parkinson's disease-associated mutations.* Journal of Neurochemistry, 2001. **77**: p. 929-34.
95. Yoshida, H., et al., *Synuclein Proteins of the Pufferfish Fugu rubripes: Sequences and Functional Characterization.* Biochemistry, 2006. **45**: p. 2599-2607.
96. Martínez-Navarrete, G.C., et al., *α-Synuclein gene expression profile in the retina of vertebrates.* Molecular Vision, 2007. **13**: p. 949-61.

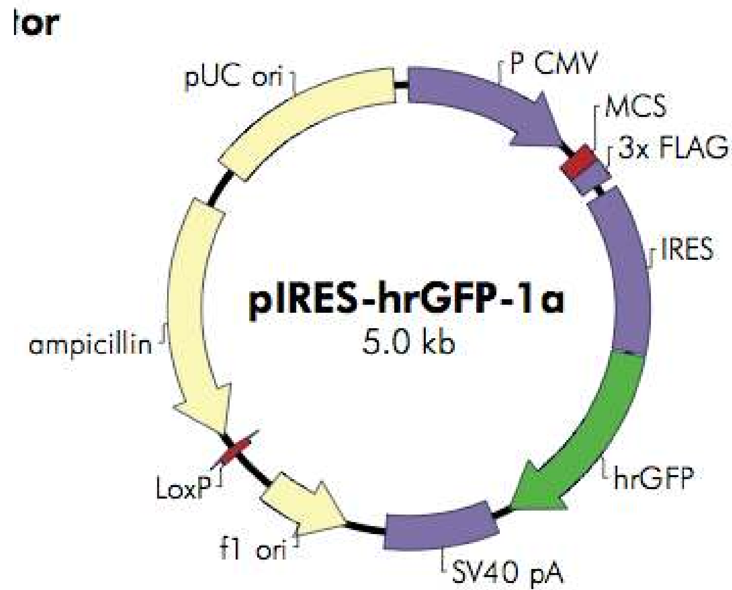
10. Annexes

Annex 1



Annex 1 –pMiniTol2 vector. This vector is specific for the Tol2 transposase technique. AmpR: Resistance for the ampicilin gene, used for selection in *E.coli*. M13F and M13R: local of recognition for M13 primers. M13 Phage is used Vectors for DNA Sequencing, Cloning and Expression Vectors,.

Annex 2



Annex 2- pIRES-hrGFP-1a. *Ampicilin*: Resistance for the ampicilin gene, used for selection in *E.coli*; *pCMV*: strong human promoter for *cytomegalovirus*; *SV40 pA*: polyadenylation signal; *IRES*: Internal Ribossome entry site; *Flag*: Flag epitope; *hrGFP*: humanized recombinant GFP; *pUC*: origin for replication and growth in *E.Coli*; *F1 ori*: origin of replication for phages; *MCS*: Multiple Cloning Site.