Beatriz Fernandes de Sousa Ribeiro

Bachelor degree in Biomedical Sciences, University of Aveiro

CRISPR/Cas9 and other molecular tools to study potentially neuroregenerative genes

Dissertation to obtain the Master's degree in

Molecular Genetics and Biomedicine

Supervisor: Sandra Isabel Moreira Pinto Vieira, Assistant Professor (DCM/iBiMED, UA)

Co-Supervisor: Margarida Castro Caldas, Assistant Professor (FCT, UNL)

Jury:

President: Alexandra R. Fernandes, Assistant Professor (FCT, UNL)

Examiner: Bibiana Isabel da Silva Ferreira, Assistant Professor (CBMR, UAlg)

Supervisor: Sandra Isabel Moreira Pinto Vieira, Assistant Professor (DCM/iBiMED, UA)

December 2020

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

NOVA School of Science and Technology

NOVA University of Lisbon

CRISPR/Cas9 and other molecular tools to study potentially neuroregenerative genes

by

Beatriz Fernandes de Sousa Ribeiro

Dissertation presented to NOVA School of Science and Technology as requirement for obtaining the Master's degree in Molecular Genetics and Biomedicine, executed under the scientific supervision of Professor Sandra Isabel Moreira Pinto Vieira, Assistant Professor in the Department of Biomedical Sciences (University of Aveiro), and Principal Investigator at the Institute of Biomedicine (iBiMED, University of Aveiro), and Professor Margarida Castro Caldas, Assistant Professor in the Department of Life Sciences (NOVA School of Science and Technology, NOVA University of Lisbon).

December 2020

CRISPR/Cas9 and other molecular tools to study potentially neuroregenerative genes

Agradecimentos

Este trabalho não se teria concretizado sem o apoio direto ou indireto de algumas pessoas a quem, por isso, não posso deixar de agradecer:

Começo por agradecer à Professora Sandra, pela orientação e apoio ao longo da realização deste trabalho, pela partilha de experiência, e pela transmissão do entusiamo pelo mundo da Ciência.

À Professora Margarida, pela partilha tão dedicada de conhecimento e por toda a disponibilidade e compreensão durante o desenvolvimento desta dissertação.

À Professora Tânia Caetano, pela disponibilidade e apoio, essenciais para ultrapassar alguns obstáculos que surgiram ao longo da realização deste trabalho.

Aos meus pais, pela presença constante, pelo amor incondicional e por todo o apoio na minha formação e na luta pelos meus sonhos e objetivos. Ao meu irmão, por acreditar em mim como ninguém.

Aos meus Amigos, pela confiança e motivação que sempre me conseguem transmitir, e pela alegria que acrescentam aos meus dias.

Ao Fábio, por acreditar (e me fazer acreditar) em mim e nas minhas capacidades, e por partilhar comigo todos os momentos, bons e menos bons, desta jornada.

Quero agradecer ao 'Zega', pelo companheirismo e pelo tanto que fomos aprendendo e crescendo juntos, bem como à restante equipa, pela disponibilidade e apoio. Em particular, agradeço à Patrícia, pelo exemplo que tem sido para mim, e pela paciência e ajuda ao longo da realização deste trabalho.

Este trabalho foi apoiado pela Fundação para a Ciência e a Tecnologia (FCT), Centro 2020 e Portugal 2020, programa COMPETE, QREN, e União Europeia (programa FEDER), através de financiamento para o projeto GoBack (PTDC/CVT-CVT/32261/2017), para o Instituto de Biomedicina (iBiMED) da Universidade de Aveiro (UID/BIM/04501/2019), e para a Plataforma Portuguesa de BioImagem (PPBI-POCI-01-0145-FEDER-022122), da qual a LiM Facility do iBiMED/UA faz parte.

Resumo

Devido à baixa capacidade neuroregenerativa dos mamíferos, a lesão vertebromedular (LVM) causa graves limitações funcionais no paciente. Não existe atualmente nenhuma terapia eficaz para esta neuropatologia, e terapias genéticas focando alvos envolvidos na fisiopatologia ou na neuroregeneração poderão ser essenciais na recuperação neurológica e funcional dos pacientes.

Num estudo bioinformático prévio do grupo, comparando três estudos de transcriptómica de roedores modelo LVM intervencionados e com melhoria locomotora, encontraram-se 68 genes diferencialmente expressos em comum. O principal objetivo desta dissertação consistiu em desenvolver ferramentas moleculares para estudar o papel de dois destes genes em processos regenerativos, o membrane palmitoylated protein 3 (*MPP3)* e o reprimo (*RPRM)*, encontrados sobre- e sub-expresso, respetivamente, no período agudo da lesão. O *RPRM* foi também encontrado sub-expresso noutro trabalho do grupo, em neurónios regenerativos.

O *MPP3* foi sub-clonado a partir de um vetor de expressão procariota para um de expressão eucariota. Após monitorização das colónias positivas por PCR, e confirmação por sequenciação, o plasmídeo recombinante foi transfetado em células de neuroblastoma SH-SY5Y, e a sobre-expressão analisada por imunocitoquímica.

Para o *RPRM*, desenhou-se um protocolo de knockout por CRISPR/Cas9, recorrendo a vetores lentivirais. Primeiramente, estabeleceram-se células SH-SY5Y-Cas9 estáveis, selecionadas com blasticidina, que serão uma ferramenta muito importante para o trabalho futuro do grupo. Clones destas células estão a ser isolados por "single-cell colony", e a expressão de *SpCas9* será confirmada por qRT-PCR. Um gRNA para edição do *RPRM* nas mesmas já foi escolhido.

Em paralelo, efetuou-se uma análise bioinformática comparativa de um dos estudos originais e um novo estudo com dados de transcriptómica de tempos pós-lesão mais posteriores, revelando novos potenciais alvos a ser posteriormente modulados por CRISPR/Cas9 nas células SH-SY5Y-Cas9.

Seguir-se-á o estudo da modulação do *MPP3* e *RPRM* na diferenciação neuronal e outros processos celulares regenerativos, relevantes para futura terapia genética em pacientes com LVM.

Palavras-chave

Lesão vertebro-medular, genes diferencialmente expressos, CRISPR/Cas9, Membrane palmitoylated protein 3 (MPP3), Reprimo (RPRM).

Abstract

Due to the low neuroregenerative capacity of mammals, spinal cord injury (SCI) causes severe functional limitations in the patient. There is currently no effective therapy for this neuropathology, and genetic therapies focusing on targets involved in SCI pathophysiology or neuroregeneration may be essential for the patients' neurological and functional recovery.

In a previous bioinformatics analysis by the group, comparing three transcriptomics studies in rodents SCI models intervened to partially increase locomotion recovery, 68 commonly differentially expressed genes were found. The main objective of this thesis was to develop molecular tools to study the role of two of these genes in regenerative processes, membrane palmitoylated protein 3 (*MPP3*) and reprimo (*RPRM*), found over- and under-expressed, respectively. *RPRM* was also found under-expressed in another work by the group, unveiling the early transcriptomic response of regenerative neurons.

MPP3 was subcloned from a prokaryotic to an eukaryotic expression vector. After monitoring positive colonies by PCR, and confirmation by sequencing, the recombinant plasmid was transfected into SH-SY5Y neuroblastoma cells, and overexpression analyzed by immunocytochemistry.

For *RPRM*, a protocol for CRISPR/Cas9 knockout was designed, using lentiviral vectors. Firstly, through selection with blasticidin, stable SH-SY5Y-Cas9 were established. These will be a very important tool for the future work of the group. Clones of these cells are being isolated by “single-cell colony” and *SpCas9* expression will be confirmed by qRT-PCR. A gRNA to edit *RPRM* in those cells has already been chosen.

In parallel, a bioinformatics analysis was performed, comparing one of the original studies and a new study with transcriptomics data from later post-injury times. This revealed new potential targets to be later modulated by CRISPR/Cas9 techniques in SH-SY5Y-Cas9 cells.

Following work will include modulation of *MPP3* and *RPRIMO* genes in neuronal differentiation and other cellular processes, potentially relevant for future gene therapy in SCI patients.

Keywords

Spinal cord injury, differentially expressed genes, CRISPR/Cas9, Membrane palmitoylated protein 3 (MPP3), Reprimo (RPRM).

Table of contents

Agradecimentos ..... *v*

Resumo ..... *vii*

Abstract ..... *ix*

Table of contents ..... *xi*

List of figures ..... *xiii*

List of tables ..... *xv*

List of abbreviations and acronyms ..... *xvii*

1. Introduction *1*

1.1. Spinal cord injury (SCI) **1**

1.2. Gene and molecular therapies for SCI **2**

1.3. Genome editing as a tool **3**

1.4. CRISPR/Cas9 system **4**

1.5. CRISPR in SCI **9**

1.6. Identifying potential neuroregenerative targets, for genetic modulation: *MPP3* and *RPRM* **10**

2. Goals *15*

3. Materials and Methods *17*

3.1. *MPP3* experimental study **17**

3.1.1. The *MPP3* plasmids 17

3.1.2. Heat shock transformation of competent bacteria 17

3.1.3. Midiprep with the NucleoBond® Xtra Midi for DNA extraction 17

3.1.4. Strategy for subcloning the *MPP3* full-length cDNA insert 18

3.1.5. Plasmids enzymatic digestion 20

3.1.6. Agarose gel electrophoresis and gel extraction 20

3.1.7. Ligation and amplification of the *MPP3* recombinant DNA 20

3.1.8. *MPP3* primers selection for positive colonies and successful cloning confirmations 20

3.1.9. Colony-Polymerase Chain Reaction (PCR) screening 21

3.2. Mammalian cells culture and manipulation **22**

3.2.1. SH-SY5Y and 293T cells culture and maintenance 22

3.2.2. Mammalian cells count 23

3.2.3. Test and compare the efficiency of different transfection reagents 23

3.2.4. *MPP3* transient transfection of SH-SY5Y cells by X-treme Gene 340 reagent 24

3.2.5. Immunocytochemistry assays and microscopy analysis 24

3.3. CRISPR/Cas9 experimental study **24**

3.3.1. Selection of the CRISPR/Cas9 system 24

3.3.2. Amplification of pLenti DNAs 27

3.3.3. Establishment of a stable neuroblastoma cell line expressing Cas9 27

3.3.4. Polybrene toxicity test and optimization of Blasticidin and Puromycin concentrations for selection of the transduced SH-SY5Y cells 29

3.3.5. Production of SpCas9-gene containing lentiviral particles 29

3.3.6. Transduction of SH-SY5Y cells with the *SpCas9* gene-containing lentivirus 30

3.3.7. SH-SY5Y Cas9 transduced cells selection (ongoing) 31

3.4. Comparative bioinformatics analysis of transcriptomic SCI studies **31**

3.4.1. Selection of transcriptomic studies 31

3.4.2. Data extraction 31

3.4.3. Identification of Differential Expressed Genes (DEGs) 32

3.4.4. Finding Human orthologs 33

3.4.5. DEGs comparisons, Gene Ontology, pathways enrichment and protein-protein interaction network analyses 33

4. Results....35

4.1. MPP3 overexpression experimental studies....35
4.1.1. Construction of a MPP3 full-length mammalian expression vector 35
4.1.2. DNA sequencing of the pBABE fl-MPP3 cDNA....38
4.1.3. Optimization of neuroblastoma cells transfection 38
4.1.4. MPP3 transient transfection of SH-SY5Y cells by X-treme Gene 340 reagent 40

4.2. *RPRM* knockout-related experimental studies....41
4.2.1. Polybrene, Blasticidin and Puromycin cytotoxicity tests for SH-SY5Y cells 41
4.2.2. Selection of SH-SY5Y cells transduced with the *SpCas9*-containing lentivirus 44

4.3. Bioinformatics analysis of the transcriptomes of partially regenerative SCI rodents....48
4.3.1. Selection of transcriptomic studies 48
4.3.2. Data extraction 50
4.3.3. Data pre-processing and DEGs screening (E-GEOD-69334 20 dpi dataset)....50
4.3.4. Conversion to human orthologs....50
4.3.5. Comparison of expression profiles 51
4.1.6. PPI network analysis 52
4.1.7. Gene Ontology and pathways enrichment analysis 53

5. Discussion 59

6. References 67

6. Annex....71

List of figures

Figure 1.1. Illustration of the pathophysiology of traumatic SCI. 2

Figure 1.2. CRISPR/Cas system in a prokaryotic cell. 5

Figure 1.3. Engineered CRISPR/Cas9 system for mammalian gene editing. 7

Figure 1.4. General Mechanism of CRISPR/Cas9 Genome Editing. 8

Figure 1.5. RPRM induces cell cycle arrest at the G2/M checkpoint. 12

Figure 3.1. Schematic protein structures of MPP3 and MPP3ΔGuK homologues. 18

Figure 3.2. Scheme of the subcloning strategy to create recombinant pBABE CMV NEO MPP3 full ..19

Figure 3.3. pLentiCas9-Blast and pLentiGuide-Puro vectors' maps. 25

Figure 3.4. Production of SpCas9-carrying lentiviral particles (LVLPs) for neuroblastoma cells transduction. 28

Figure 3.5. Timeline of the experiments regarding the establishment of a stable neuroblastoma cell line expressing SpCas9. 29

Figure 4.1. Scheme of expected results for the enzymatic restriction of the MPP3 plasmids. 35

Figure 4.2. Results obtained for the enzymatic restriction of the MPP3 plasmids. 36

Figure 4.3. Gel electrophoresis with the results from the colony-PCR screening for pBABE fl-MPP3 transformants. 37

Figure 4.4. Confirmation of the pBABE fl-MPP3 DNA sequence. 38

Figure 4.5. Scheme of the transfection conditions tested in a 6-well plate. 38

Figure 4.6. Epifluorescence microphotographs for qualitative comparison of the efficiency of three different transfection reagents. 39

Figure 4.7. Epifluorescence microphotographs of the transient transfection of SH-SY5Y cells with pBABE fl-MPP3. 40

Figure 4.8. Phase contrast microphotographs of a blasticidin dose-response experiment in SH-SY5Y cells, exposed for up to 5 days to 2.5, 5.0 and 10 µg/mL of blasticidin. 42

Figure 4.9. Phase contrast microphotographs of a puromycin dose-response experiment in SH-SY5Y cells, exposed for up to 5 days to 1, 2 and 4 µg/mL of puromycin. 43

Figure 4.10. Selection of positively transduced SH-SY5Y-Cas9 neuroblastoma cells with 5 µg/mL of blasticidin for 7 days. 45

Figure 4.11. Phase contrast microphotograph of transduced SH-SY5Y cells, on the 6th day of selection with 5 µg/mL blasticidin. 46

Figure 4.12. Long-term selection of SH-SY5Y-Cas9 cells at day 25. 47

Table 4.2. Number of identified up-regulated (UR) and down-regulated (DR) DEGs. 50

Figure 4.13. Comparison of expression profiles of the two analyzed studies. 51

Figure 4.14. Protein interactions between up-regulated common genes. 52

Supplementary Figure 1. Alignment between the complete sequencing readout of the selected positive clone DNA, and the pBABE fl-MPP3 and pBABE t-MPP3 sequences. 73

List of tables

Table 3.1. Components of the enzymatic restriction reactions for MPP3 subcloning. 20

Table 3.2. Components of the colony-PCR reaction for MPP3 positive colonies screening. 21

Table 3.3. Colony-PCR conditions, for MPP3 positive colonies screening. 22

Table 3.4. Characteristics of available gRNAs for human RPRM knockout. 26

Table 4.1. Information and characteristics of the selected transcriptomic studies of rodent spinal cord injury models with regenerative approaches. 49

Table 4.2. Number of identified up-regulated (UR) and down-regulated (DR) DEGs, with p-adjust<0.05 and a |log2FC| ≥ 1, found in the E-GEOD-69334 and PMC6854783 studies. 50

Table 4.3. Number and percentage of UR and DR DEGs with p-adjust<0.05 and a |log2FC| ≥ 1, for which human orthologs were found, per study. 51

Table 4.4. List of the common up-regulated genes between E-GEOD-69334 and PMC6854783. 54

Supplementary Table1. Studies on potential therapies for SCI that include a transcriptomic analysis. 74

List of abbreviations and acronyms

AdV – Adenovirus

AAV – Adeno-associated virus

AuNP – Inorganic gold nanoparticle

Bp – Base pairs

CDS – Coding Sequence

circRNA – Circular RNA

CNS – Central Nervous System

CPP – Cell-penetrating peptide

CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats

DBS – Double stranded DNA break

DEG – Differentially regulated gene

ECM – Extracellular matrix

HDR – Homology-directed repair

lncRNA – Long non-coding RNA

LVLPs – Lentivirus like particles

MAGUK – Membrane-associated guanylate kinase

miRNA – Micro RNA

MPP3 – Membrane Palmitoylated Protein 3

MPSS – Methylprednisolone Sodium Succinate

NASCIS – National Acute Spinal Cord Injury Study

NF – Neurotrophic factor

SCI – Spinal Cord Injury

PNI – Peripheral Nerve Injury

PNS – Peripheral Nervous System

RAG – Regeneration Associated Gene

RPRM - Reprimo

TALE – Transcription activator-like

TALEN – Transcription activator-like effector nucleases

ZFN – Zinc Finger Nucleases

1. Introduction

1.1. Spinal cord injury (SCI)

Spinal cord injury (SCI) refers to the insult of the spinal cord, most commonly of traumatic nature, caused by a fracture or dislocation of the vertebrae[1,2]. The fracture and dislocation of vertebral bodies, lead to the compression of the spinal parenchyma, and, consequently, to variable degrees of permanent neurological dysfunction, depending on the degree of damage and the spinal level at which the injury occurs. Neurological impairments include motor, sensory and autonomic (sexual, urinary, cardiovascular and intestinal) dysfunctions[1,3,4]. Being highly debilitating, SCI brings heavy emotional and economic burdens to individuals, families, and societies, being increasingly recognized as an important worldwide health priority[3,5]. In the world, there are approximately 2 million people living with SCI, and 700,000 new cases per year[6]. In Portugal, the annual incidence is approximately 58 cases per million individuals[7].

There is a need to improve survival, as well as the quality of life for these patients[8]. However, despite the global research intense efforts to understand the pathophysiological mechanisms underlying this Central Nervous System (CNS) pathology, there are currently no effective and reliable clinical treatments for this condition[8]. Therefore, the medical management of individuals with SCI throughout their lifetime remains challenging[8].

The pathophysiological processes of SCI can be categorized into primary and secondary injury (or lesion) (Figure 1.1). The primary injury is characterized by the contusion and compression of the spinal cord that results in axonal and cellular damage at the site of the injury. This initial mechanical injury leads to a secondary injury, characterized by a cascade of biochemical events, including ischemia, edema, inflammation, cellular necrosis, free radical formation, lipid peroxidation, changes in ionic balance, glutamate excitotoxicity and vascular changes, within and around the primary lesion site[2]. Moreover, resident microglia, as well as neutrophils, macrophages and lymphocytes recruited from the periphery (since the blood-CNS barrier is broken), are activated[9]. These cells, as well as reactive astrocytes, secrete a range of factors that lead to the resolution of the acute injury response, wound healing, and to scar and cavity formation, which may also amplify the inflammatory response and contribute to the inhibition of axonal regeneration[10]. Hence, although mild inflammation can have positive effects following a CNS lesion, such as to activate microglia cells to promote differentiation of neural precursor cells into neurons, high levels of inflammatory mediators lead to deleterious effects such as neurotoxicity, neurodegeneration and, consequently, aggravation of tissue damage[9]. Excessive inflammation also impedes neuroplasticity, which is essential to the recovery of function, as it potentiates the sprout and regeneration of spared and injured axons, increasing the strength of residual connections, and also promotes the formation of new connections and circuits[1,8].

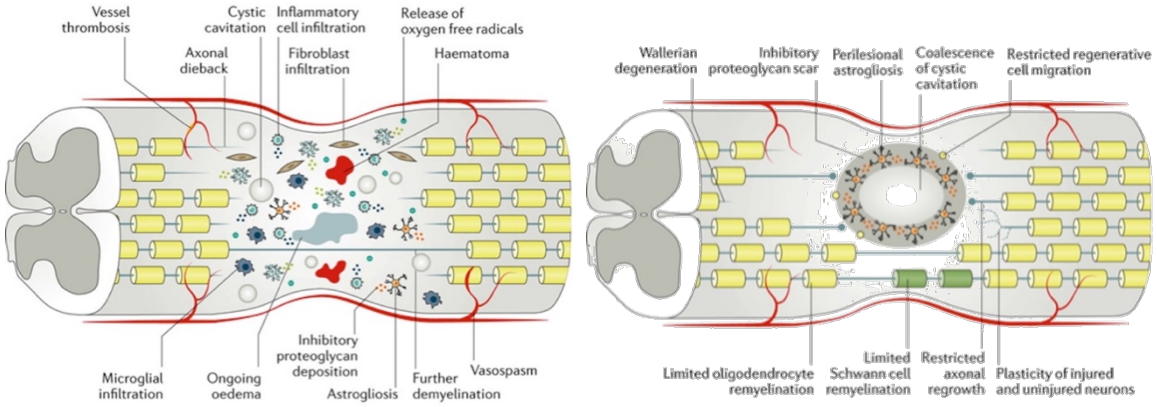


Figure 1.1. Illustration of the pathophysiology of traumatic SCI. The initial trauma to the spinal cord is followed by a secondary injury characterized by inflammatory cell infiltration, ischemia, demyelination and others, leading to the formation of the inhibitory glial scar. Adapted from Badhiwala, J. (2019)[11].

At the clinical level, the current standard procedure in the first hours after trauma involves the administration of high doses of corticosteroids (methylprednisolone sodium succinate, MPSS), potent anti-inflammatory drugs, aiming to improve neurological outcomes[10,12]. However, there is still a lack of evidence supporting its clinical benefit[1,3,12,13]. Although the National Acute Spinal Cord Injury Study (NASCIS) has reported improved motor and sensory scores at the follow-up rehabilitation period after treatment with high MPSS doses, some criticism has arisen against these trials, including the studies' designs, data quality, statistical analyses, interpretation, and conclusions[12]. Additionally, some studies demonstrated associations between the early use of high MPSS doses in acute SCI and increased rates of gastrointestinal hemorrhage and wound infections, making its use controversial[10,14]. Other therapeutics for SCI that have been proposed and are currently under investigation mainly focus on three areas: neuronal protection, to prevent harm due to secondary injury processes, such as inflammation and cell death; neuronal regeneration to promote axonal regrowth, guidance and synapse formation, repairing injured neuronal pathways and eventually leading to the recovery of motor and sensory functions; and tissue engineering technologies, to help in tissue reconstruction, which may involve only biomaterials or their use to deliver cells, genes or proteins to the injured tissue, providing a regenerative route for axonal regrowth[15].

1.2. Gene and molecular therapies for SCI

The low supply of neurotrophic factors (NFs) to the injury site difficults the promotion of neuronal regeneration[16]. Additionally, the presence of inhibitory factors such as axonal growth inhibitors (AGIs) like Nogo receptor-1 (NgR1) and its five ligands, in the extracellular matrix (ECM), also contributes for the limited regenerative capacity following SCI[16,17]. Therefore, SCI frequently results in permanent neurological and functional deficits in affected individuals. Indeed, the neurons of the CNS do not spontaneously regenerate their axons following injury, contrary to the

dorsal root ganglion (DRG) neurons of the Peripheral Nervous System (PNS), in which regeneration associated genes (RAGs) are activated[18]. Manipulation of some of these molecules and RAGs in the injured CNS neurons may help improving repair.

To find current gene and molecular therapies for SCI, a search on the Pubmed database was made, of studies published until March of 2020, using the keywords “spinal cord injury gene therapy” OR “spinal cord injury molecular therapy” OR “central nervous system regeneration therapy”, filtered to include reviews in this topic and also current *in vitro* and *in vivo* studies.

In some of the retrieved studies, researchers have demonstrated that inhibiting AGIs in SCI animal models leads to an increase in anatomical plasticity and axonal regrowth of functionally important nerve fiber systems, in this way promoting functional recovery[15]. For example, the use of chondroitinase ABC and anti-NogoA treatments to reduce inhibitory signaling in the neuronal extrinsic environment after SCI, have shown promising results in terms of promoting axon sprouting and recovery[1,19].

The contrast between the CNS and the PNS transcriptional responses to injury, as well as the complexity of the secondary injury over time, sparked the research on the temporal and spatial changes in gene expression and molecular regulation following SCI (and PNI). An increasing effort has been made to identify genes and signaling pathways that could be manipulated to, eventually, minimize the extent of the lesion and/or improve recovery, through the promotion of neuronal growth capacity in the injured CNS[18,20].

Although single-gene manipulation, in experimental PNS and CNS injury models, allowed to enlighten several genes and signaling pathways whose manipulation interferes with axon regeneration, this last is a process in which broad genetic programs are involved[18]. As such, efforts are being made to identify a smaller number of upstream molecular regulators of transcriptional programs that drive the regeneration in the PNS and, importantly, if possible, to recapitulate these programs in CNS neurons[18].

Transcriptomics are thus being used in the search for biological targets to promote neural repair, and not only focus on expressed genes but also on other molecular factors, such as noncoding RNAs (ncRNAs). Epigenetic and post translational modifications modulators, small molecules as ligands and other signaling modulators, are also focus of study[18]. Bioinformatic tools are an important asset to this type of investigation, enabling the identification of upstream transcriptional regulators, the correlation of RNA-Seq and protein expression data, and the prediction of small molecules that can activate the expression of genetic programs[3,18].

1.3. Genome editing as a tool

Genome/gene editing, or engineering, consists in the alteration, in a specific and targeted manner at particular locations in the genome, of an organism’s DNA sequence, including addition, deletion, or modification, based on the natural repair mechanisms of the cell when facing double stranded DNA breaks (DBSs)[21,22].

DSBs in eukaryotes are repaired by one of two endogenous repair mechanisms: non-homologous end joining (NHEJ) or homology-directed repair (HDR). NHEJ can occur at any phase of the cell cycle and is the primary and most common cellular DSB repair mechanism, where protein factors re-ligate the broken DNA strand either directly or by including nucleotide insertions or deletions ('indels'). This process is error-prone, since it occurs without a homologous DNA template and frequently leads to mutations and deletions in the repaired strand. In contrast, HDR uses a homologous repair template to precisely repair the DSB, so this mechanism is less susceptible to errors. HDR typically occurs in late S- or G2-phase, when a sister chromatid can serve as the repair template. Therefore, in general, the incidence of HDR for DSB repair is extremely low when compared with NHEJ, at least in instances where both pathways are equally available to an organism. These mechanisms can be exploited to induce gene knockouts or knock-ins, as it will be discussed later[23].

The breakthrough in genome engineering came with the design of nucleases that were able to specifically recognize and cleave target DNA, activating the repair pathways and generating mutations. The current state-of-the-art technique for gene edition is CRISPR/Cas (for Clustered Regularly-Interspaced Short Palindromic Repeats, and CRISPR associated protein), further discussed below. Before the advent of CRISPR/Cas9, the two approaches more commonly used to make site-specific DBSs and gene editing were based on zinc finger nucleases (ZFNs) or on transcription activator-like effector (TALE) nucleases (TALENs)[21,22]. The ZFNs system is based on two ZFNs, engineered to recognize different closely located nucleotide sequences within the target site. It requires the simultaneous recognition and binding of both ZFNs, which limits off-target effects. Zinc finger domains can recognize a trinucleotide DNA sequence and, although a series of linked zinc finger domains can recognize longer DNA sequences, zinc finger motifs influence the specificity of neighboring zinc fingers, making the design and selection of modified zinc finger arrays more challenging and more time-consuming. In the case of TALENs, a single TALE motif recognizes one nucleotide and an array of TALEs can associate with a longer sequence, without the activity of each TALE domain affecting the binding specificity of neighboring TALEs, which makes the engineering of TALENs much easier than ZFNs. In both these systems, ZFNs and TALE motifs are linked with FokI endonuclease, which requires dimerization for the cleavage to occur. This means that the binding of two different ZFNs/TALENs at opposite strands in close vicinity to the target DNA is also needed[22].

1.4. CRISPR/Cas9 system

In 2012, the scientists Emmanuelle Charpentier and Jennifer A. Doudna adapted the CRISPR/Cas9 system into a genetic editing tool[18], revolutionizing the field of molecular biology, gene therapy, and precision medicine, being worthy of the Nobel Prize in Chemistry in 2020. It has been mostly used to edit the genome of cells in culture and develop animal models, to understand disease mechanisms and validate treatment targets[22,24].

The CRISPR/Cas system is actually a component in the acquired/adaptive immune system of bacteria or archaea, aimed at destroying (cutting and degrading) exogenous nucleic acids such as DNA of invading bacteriophages or plasmids from other organisms (Figure 1.2)[25]. It keeps a record of previous infections, in order to reduce the effects of a future infection by the same agent[26]. Contrary to the mammalian immune system, it does not use antigens and antibodies to recognize and record a specific infection. Instead, it recognizes the invasive nucleic acid and copies a portion of it – segments called 'spacers' that range from 26 to 72 base pairs (bp) – and incorporates it between repeat sequences, that also vary in size, from 21 to 42 bp.

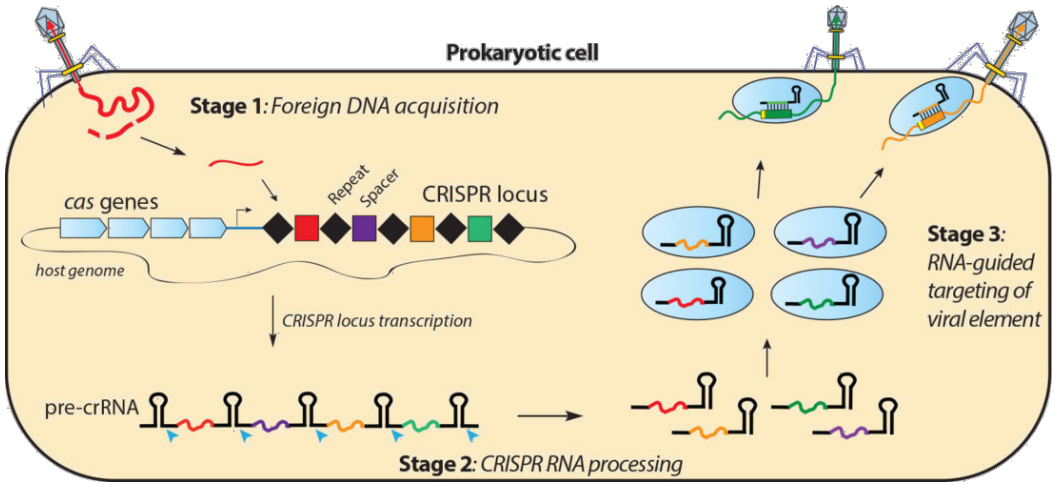


Figure 1.2. CRISPR/Cas system in a prokaryotic cell. The host genome keeps a record of previous infections by copying portions of the foreign DNA and incorporating them into the CRISPR locus (spacers), between repeat sequences. Upon further invasions, the target nucleic acid sequence is recognized through the base pairing with crRNAs. The noncoding trans-activating RNA (tracrRNA) mediates the attachment of Cas nuclease to the crRNA, forming a tracrRNA:crRNA:Cas9 complex able to target and destroy the viral element. Adapted from Doudna lab webpage (https://doudnalab.org/research_areas/crispr-systems/), accessed in November, 2020.

The CRISPR system is not the same among all bacteria and archaea and it can be classified into two classes (class 1 and class 2), generally encompassing 3 different types and several subtypes, according to their characteristics[22]. Class 2 is the one characterized by requiring only one protein, like Cas9, to scan, bind and cleave the target DNA sequence. Owing to the simplicity of the Type II (class 2) system, researchers first exploited the use of CRISPR for gene editing using the Cas9 system from *Streptococcus pyogenes* (SpCas9). Besides Cas9 nuclease, this system is composed by two types of RNAs, to discriminate between self and non-self DNA, which is accomplished by screening parts of the exogenous DNA with reference to endogenous RNA from the internal CRISPR locus (CRISPR RNAs or crRNAs). The recognition of the invading nucleic acid sequence occurs through the base pairing of the target with crRNAs, mediated by noncoding trans-activating RNA (tracrRNA) to form a tracrRNA:crRNA:Cas9 complex. crRNA is a 17-20 nucleotide sequence complementary to the target DNA (depends on the target gene). In contrast, the tracrRNA is an invariable sequence that serves as a scaffold to attach the Cas nuclease to the crRNA. Following transcription, the tracrRNA and pre-crRNA are stabilized by Cas9 and base

pair, and RNase III processes the pre-crRNA into crRNA by cleaving it at the repeat. This dependence on RNase III likely explains why Type II systems are found in bacteria and not archaea, as RNase III is not found in archaea. The crRNA:tracrRNA:Cas9 complex forms the active crRNA-guided endonuclease.

In 2012, when the CRISPR/Cas9 system of *S. pyogenes* was, for the first time, used as RNA-guided DNA endonuclease *in vitro*, to introduce DSBs in a target DNA, it was also demonstrated that the duo-tracrRNA:crRNA units could be engineered into a single, truncated RNA chimera and still direct efficient DNA cleavage[26]. This further simplified CRISPR/Cas9 into a two-component system: a bacterially-derived nuclease such as Cas9, and a single guide RNA (sgRNA/gRNA) (Figure 1.3.). The gRNA is thus a specific RNA sequence designed to recognize and direct the nuclease to the target genomic location. It consists of a combination of the crRNA and the tracrRNA[22].

In comparison with previous genome editing systems, while both ZFNs and TALENs have their target specificity from protein–DNA association, recognition of the DNA site in the CRISPR/Cas system is controlled by RNA–DNA interactions, through Watson-Crick base pairing[22]. Therefore, CRISPR/Cas avoids the need for protein engineering to develop a site-specific nuclease against a specific DNA target sequence, requiring only the synthesis of a RNA oligonucleotide, simplifying and reducing the time needed for gene editing design and implementation. Also, it allows the modification of several genomic sites simultaneously (multiplexing). Compared to ZFNs and TALENs, the CRISPR/Cas9 system is simpler, cheaper, and more efficient[26].

For successful binding of the CRISPR complex to the DNA, a protospacer adjacent motif (PAM) must be present downstream of the target site. PAMs are short sequences between 3 and 8 nucleotides in length, and which varies depending on the nuclease used, that is, nucleases isolated from different species require different PAMs. Most Cas nucleases recognize multiple PAM sequences[27]. For the most commonly used nuclease, SpCas9, the preferably PAM sequence is 5'-NGG-3' (where 'N' is any nucleotide). Upon recognition of the PAM sequence, the Cas9:RNA complex unwinds the DNA from the first 10–12 nucleotides upstream the PAM sequence, termed the 'seed region'[22,27]. If the interrogated DNA sequence matches the crRNA target sequence, the Cas9 HNH nuclease domain cleaves the target strand complementary to the gRNA, while the RuvC nuclease domain cuts the non-complimentary strand, together creating a DSB that occurs three nucleotides upstream of the PAM[22,26]. Although the requirement of a PAM motif adjacent to the 20 nucleotide long target sequence complementary to the guide RNA might be a limitation for SpCas9, 5'-NGG-3' sequences have been reported to occur every eight bps, on an average, in the human genome, which represents an advantage for its use[22,28].

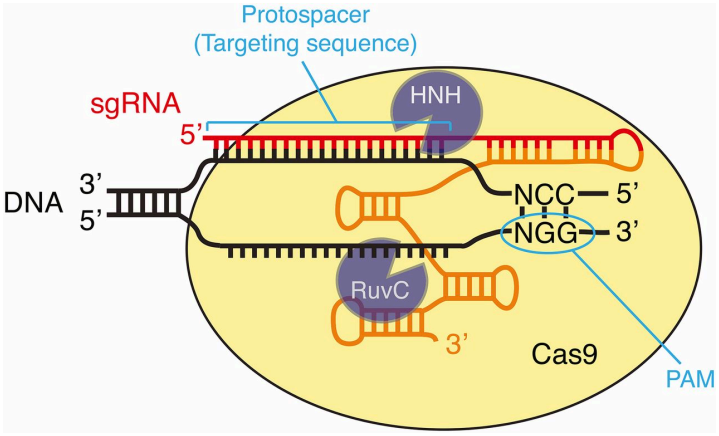


Figure 1.3. Engineered CRISPR/Cas9 system for mammalian gene editing. Adapted from Dickinson, D (2016)[25].

Once the nuclease cleaves the DNA, the native cellular DNA repair machinery attempts to repair the DSB through one of the two mechanisms above described (NHEJ or HDR; Figure 1.4.). NHEJ, the primary cellular DSB repair mechanism, re-ligates the DNA ends in the absence of a homologous DNA template, introducing indels into the target DNA. Indels resulting in frameshift mutations and premature stop codons can produce a loss-of-function (LOF) mutation, and this is the primary means by which CRISPR is used to disrupt (knockout) a gene. HDR can be performed to replace and later express a specific genetic sequence (knock-in). Hence, in addition to the main CRISPR components, HDR-mediated CRISPR editing requires a DNA donor template containing the new desired sequence flanked by regions of homology[25]. Introduction of this donor template, along with the CRISPR components, allows the cells to repair the DSB via homologous recombination, incorporating the new sequence into the target gene. Noteworthy, NHEJ is still the more frequent repair mechanism observed from CRISPR/Cas9 editing, even in the presence of a donor template DNA[29,30].

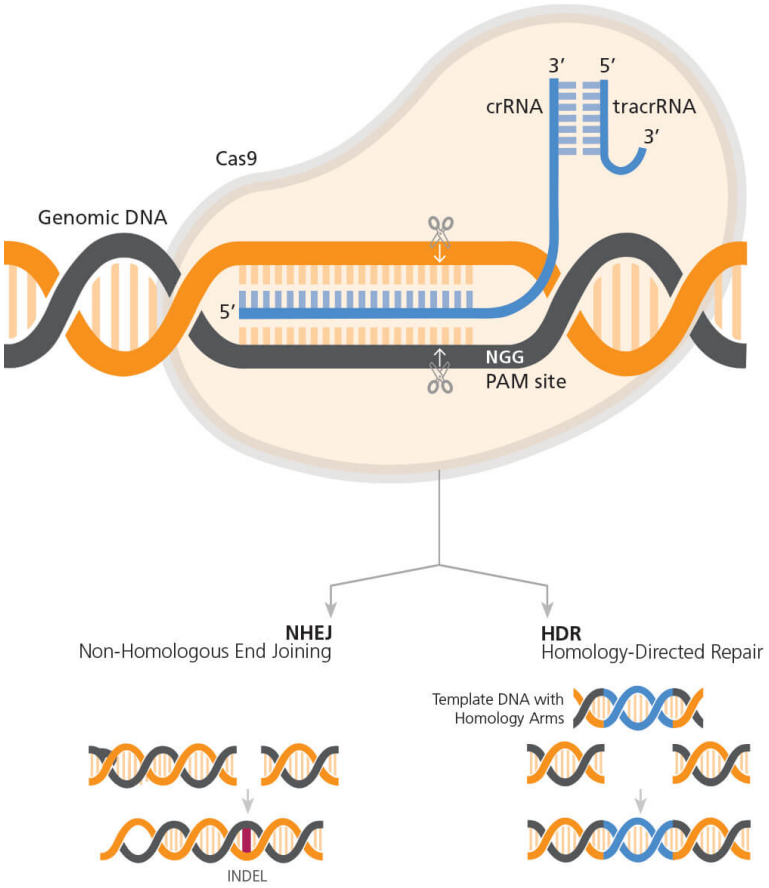


Figure 1.4. General Mechanism of CRISPR/Cas9 Genome Editing. After Cas9 generates a break in the target DNA sequence, it triggers one of the two endogenous DNA repair pathways: while non-homologous end joining (NHEJ) can result in indels and consequent knockout of the gene, homology-directed repair can use a template DNA to produce a knock-in of a precise sequence. Adapted from Stem Cell Technologies website (https://www.stemcell.com/genome-editing-of-human-primary-t-cells-using-the-arcitect-CRISPR/Cas9-system.html), accessed in November, 2020.

Single mismatches, and sometimes multiple mismatches, can be tolerated by the Cas9:RNA complex, with mismatches being more tolerated in regions downstream the seed region. Nonetheless, it is important to refer that some studies have shown that SpCas9 can sometimes induce unintentional mutations in the genomic loci that are different from the target sites[22].

In terms of the delivery of CRISP components into cells, there are multiple ways to do it. They can be supplied via physical administration, including microinjection, electroporation and hydrodynamic deliver (currently under investigation); via lipid nanoparticles, liposomes, lipoplexes/polyplexes, cell-penetrating peptides (CPPs), etc. CRISPR components can be delivered through plasmid DNA, which can be transfected into cells using the majority of available transfection methods, in a simple and cost-effective delivery method. With this method, the nuclease and gRNA can be encoded together in a single DNA plasmid or separately in two individual plasmids, which is usually more efficient. Cas9 can also be directly delivered as a protein. Another common and effective way to deliver these components is via viral vectors, including specifically engineered adeno-associated virus (AAV), lentivirus (LV), or adenovirus (AdV). Although non-viral vector delivery carries lower risk of immune responses than viral ones,

non-viral vectors have a poor transduction efficacy and low potential to deliver CRISPR/Cas9 in clinical settings, at least in the near future[24].

More recently, new delivery technologies have been introduced such as multifunctional envelope-type nano-devices, lipid-coated mesoporous silica particles, and inorganic gold nanoparticle (AuNPs) incorporating Cas9, gRNA and donor DNA[22].

1.5. CRISPR in SCI

As referred before, gene therapy represents a promising approach to enhance spinal cord recovery after injury. However, it will be important to direct some of these molecular/genetic approaches to target cells discriminately, glial or neuronal, since characteristic changes in the microenvironment of the injured spinal cord lead to unique transcriptomic changes[31,32].

Thus far, preclinical investigations reveal the applications of *in vivo* gene therapy in the injured spinal cord to: 1) enhance the expression of pro-regenerative factors, 2) molecularly modulate neural circuits, 3) block the expression of detrimental proteins, and 4) introduce matrix-modifying enzymes for the degradation of inhibitory particles[33].

Due to its high versatility, specificity, efficiency, and also, ease of use and affordable costs, CRISPR/Cas9 is one of the most powerful tools for editing genomes and it has already been applied to develop therapeutic strategies for cancer and many other genetic diseases[31,32]. Regarding its application in the CNS, as potential therapy, it has been tested in post-mitotic neuronal cells, showing some limitations, such as related with the direct delivery of the CRISPR/Cas9 machinery in the form of Cas9 protein-gRNA ribonucleoprotein (RNP), since it has shown better results of gene editing efficiency, comparing to the encoding plasmids[31,34]. Although the use of viral vectors is more common, neuronal transduction using those techniques has not being much reported for gene modulation[31]. Despite that, recently, a dual lentiviral CRISPR-based transcriptional activation (CRISPRa) system has shown high efficiency *in vitro* and *in vivo* transformation of neurons and post-mitotic neurons, respectively[31].

The CRISPR/Cas9 system has been tested *in vitro* and *in vivo*, in pre-clinical models of SCI, such as the axolotl. This animal has the ability to completely regenerate its spinal cord, structurally and functionally, due to its ependymal cells in the central canal, that remain Neural Stem Cells (NSCs)[34]. Understanding the molecular mechanisms behind the pluripotency of the NSCs and its activation upon injury, might help in the development of therapeutic strategies for human patients, and the CRISPR/Cas9 system was used in the SCI axolotl model, to target and knockout genes in NSCs at specific regions of the spinal cord and timepoints post-injury[34].

Another study, using a different animal model (mice), used the CRISPR/Cas9 technology to knock-in a transgenic allele of *TFE3*, and test if its overexpression promoted autophagic flux, potentially important for neuronal survival[35]. The authors could not confirm this, and supposed some off-target mutations occurred.

Other study[36] used CRISPR/Cas9 in rat mesenchymal stem cells to test if the modulation of certain genes could increase the anti-oxidative capacity of those cells, and help them surviving in the oxidant micro-environment of the injured regions, aiming to e.g. increase the viability of transplanted cells for therapy.

1.6. Identifying potential neuroregenerative targets, for genetic modulation: *MPP3* and *RPRM*

In our group, a comparative analysis of transcriptional data from manipulated SCI rodent models with improved functional recovery was previously performed by Correia PD et al[37]. This consisted in a bioinformatic analysis of 3 transcriptomic studies of rodent models of SCI, where 1) a regenerative approach was applied with at least partial success in terms of functional recovery, and 2) a transcriptomic study of the expression pattern of all genome at the lesion site was performed. The main goal of this comparative analysis was to identify common differentially expressed genes (DEGs) between these studies, which could represent putative regeneration associated genes (RAGs). The three studies included were: 1) E-GEOD-69334[38]; 2) E-GEOD-34430[39] and 3) GSE62760[40]. Although the selected studies presented different types of spinal cord lesion, as well as different regenerative treatments attempts, and two different rodent species (*Mus musculus* and *Rattus norvergicus)*, they shared similar lesion sites and all affirmed that the regenerative genetic or scaffold plus chemical strategies have improved recovery after lesion, in the treated animals.

From this previous work and from an ongoing comparative transcriptomic study of the early responses to injury of the non-regenerative CNS neurons *vs* PNS regenerative ones, that uses RNA-sequencing of neurons from SCI and PNI rat models, some putative RAGs arose. From these, two genes were selected to perform further studies in the context of this thesis, the *membrane palmitoylated protein 3* (*MPP3)* gene, which is up-regulated in the three studies comprehending the first comparative analysis above mentioned, and reprimo (*RPRM*), that is down-regulated during the first week after injury in those three studies, and also found down-regulated in the second, ongoing, PNI vs SCI comparative study. These are the molecular targets to which molecular tools will be developed for their modulation, aiming to study their cellular functions potentially relevant for neuroregeneration.

Membrane palmitoylated protein 3*, MPP3*. Also known as *DLG3* (discs large homolog 3) and *SAP102* (synapse-associated protein 102), *MPP3* is located at the band 21.31 of the long arm of the chromosome 17 (17q21.31), and spans 32.38 kb of genomic DNA. It belongs to a family of membrane-associated proteins termed 'MAGUK' (membrane-associated guanylate kinase homologs), whose members are localized at the membrane-cytoskeleton interface of cell-cell junctions[41,42].

MAGUK proteins appear to have both structural as well as signaling roles, regulating intracellular junctions and epithelial cell polarity. Additionally, these proteins play an important role at synaptic

junctions, where they seem to regulate synaptic development and plasticity, and the release of neurotransmitters from synaptic vesicles[41,42]. In the particular case of the MPP3 protein, it is known that it plays important roles in different polarized cell types and in the establishment and maintenance of apical cell junctions and tight junctions of epithelial cells and neuronal synapses, contributing for their development[42,43].

According to the GTEx database (https://www.gtexportal.org/), *MPP3* mRNA is overexpressed in brain, particularly in the cerebellar hemisphere and cerebellum, and in heart, in the atrial appendage and left ventricle. Moreover, the level of *MPP3* mRNA is greater in cancerous breast, kidney, liver, lung, and ovarian tissues than in normal tissues[44]. According to LifeMap Discovery – Embryonic Development & Stem Cell Compendium (https://discovery.lifemapsc.com), the *MPP3* mRNA expression is up-regulated in adult dopaminergic neurons. The MPP3 protein seems to be highly expressed in fetal brain, frontal cortex, spinal cord, and fetal heart according to ProteomicsDB (https://www.proteomicsdb.org), PaxDb (http://pax-db.org/), and MOPED – Multi-Omics Profilling Expression databases (https://www.proteinspire.org/MOPED/).

In the above-mentioned bioinformatic comparative study[37] on SCI partially regenerative rodent models, *MPP3* was found up-regulated at 3, 4 and 7 days post-injury (log2FC values of 0.31, 0.30, and 0.53, respectively), returning to similar levels as the ones of non-regenerative injured rodents from 10 days post-injury onwards[37]. Given these initial time points, where synaptogenesis is thought to be absent or relatively sparse, and MPP3 known role in cytoskeleton-membrane connection and acquisition of cell polarity, we hypothesize that MPP3 can have an early positive influence in neuritogenesis, the formation of neurites that precedes the formation and maturation of synapses, and that must occur at the first regenerative phases.

Reprimo (*RPRM*). The other gene that emerged as a common regulated gene in all spinal cord partial regenerative rodent paradigms studied by our group, was *RPRM*, found down-regulated during the first week after injury. This gene was also chosen to be further studied because it i) showed an interesting log2FC value during the first week post injury: at 3, 4 and 7 days post injury, its log2FC was of -0.39, -0.57 and -2,26, respectively; after this, its log2FC becomes positive (from 10 to 90 days post-lesion); ii) it was also found early down-regulated in *Xenopus tropicalis* (a species that can regenerate its spinal cord when at the tadpole stage) at 6, 24 and 60 hours after injury[45]; iii) it was found down-regulated at 24h post-injury in regenerating PNS neurons (our ongoing work). This may reflect the importance of this gene during the acute phase of SCI, a critical phase to act to try to reduce the degree of tissue damage. Additionally, transcriptome studies in several vertebrate species reveal that *RPRM* genes are mainly expressed in the central nervous system, including in the spinal cord[46].

The *Reprimo* gene family is a group of single exon and intronless genes – *RPRM*, *RPRML* and *RPRM3* – present exclusively within the vertebrate lineage. In particularly, only *RPRM* and *RPRML* are expressed in most of the vertebrate lineages, including human. These three genes are associated with the developmental patterning of the gastrointestinal tract, brain and blood vessels[46].

RPRM is located at the band 23 of the long arm of the chromosome 2 (2q23) and spans 1.47 kb of genomic DNA. Its product is a highly glycosylated cytoplasmic protein, with 109 amino acids and with an estimated molecular weight of 11.774 Da, found predominantly in the cytoplasm. It has two *N*-glycosylation sites, at amino acids 7 and 18, a potential serine-phosphorylation site at residue 98, a predicted sumoylation site at position 82 and a potential transmembrane domain covering amino acids 56 to 76[46,47]. In different vertebrates, its expression seems to be specific and conserved in the brain and specific sensory organs, meaning it could play an important role during sensory neurons development[48].

Functional analyses have suggested that *RPRM* is a transcriptional target for the tumor suppressor gene p53, involved in cell cycle arrest at the G2/M checkpoint (Figure 1.5)[49]. When DNA is damaged, the G2 checkpoint blocks the entry of the cell into mitosis. P53 regulates the G2/M transition through the inhibition of the cyclin-dependent kinase Cdc2, essential for mitosis to happen[50]. To be activated and allow the cell to enter mitosis, Cdc2 requires the biding to Cyclin B1 and to be phosphorylated at threonine 161 by CDK-activating kinase. The Cdc2/Cyclin B complex is kept inactive during G2 by Cdc2 phosphorylation on tyrosine 15 and threonine 14 by the kinases Wee1 and Myt1, respectively. However, at the onset of mitosis, the phosphatase Cdc25 dephosphorylates both these residues. The Cdc2/Cyclin B complex can then phosphorylate Cdc25, further activating it and initiating a positive feedback loop[50].

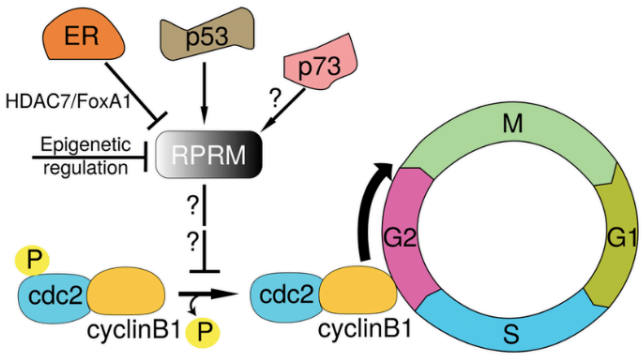


Figure 1.5. RPRM induces cell cycle arrest at the G2/M checkpoint. *RPRM* is a transcriptional target for the tumor suppressor gene p53, inhibiting the activation of the Cdc2/Cyclin B1 complex, essential for the cell to enter mitosis. Adapted from: Amigo, J[46].

The translocation of (Cdc2-)cyclin B1 into the nucleus is one of the hallmarks of cells entering the M phase[51]. p53 has various mechanisms by which it inactivates Cdc2, as repression of the *Cdc2* and the *Cyclin B1* genes. This later results in loss of Cdc2 activity and also in a decreased extent to which Cdc2 is tyrosine phosphorylated, since the Cdc2/Cyclin B1 complex is a better substrate for the tyrosine 15 kinase Wee1 than is Cdc2 alone[50]. In response to genotoxic stress, activated p53 also induces transcription of the *rprm*, *B99*, and *mcg10* genes. *RPRM* expression seems to inhibit the dephosphorylation of Cdc2 that keeps it inactive, consequently suppressing the activation of the Cdc2-Cyclin B1 complex. This suggests a potential role for RPRM as a tumor suppressor gene. Accordingly, loss of *RPRM* expression has been reported in gastric cancer and other non-digestive tumors including breast, renal cell carcinoma, adrenocortical and pituitary

tumors. Epigenetic silencing, mainly by DNA methylation, of the promoter region of *RPRM*, occurs at early stages of human cancer[47,49]. Actually, *RPRM* is located within a CpG-enriched region. These regions often cluster near gene promoters (CpG islands), thus controlling local chromatin structure and transcription factor binding. An aberrant pattern of DNA methylation, namely a local hypermethylated state in the promoter regions of tumor suppressor genes, may result in gene silencing, and appears to occur in the case of *RPRM*[46]. Nonetheless, enhanced *RPRM* expression in metastatic brain tumors has also been described[46]. Noteworthy, the molecular mechanisms and signaling partners by which RPRM exerts its p53-dependent cell cycle arrest mediator functions remain unknown[46,47,51].

Regarding our neuroregenerative studies, two questions arise. In the transcriptomic studies of spinal cord partially regenerating post-injury tissues[37], is *RPRM* down-regulated in supportive mitotic cells to allow for cell proliferation and to increase cell survival? Moreover, and most importantly, why is RPRM down-regulated in post-mitotic neuronal bodies of regenerative PNS neurons immediately at 24h post-lesion (our ongoing work), and also potentially in the neurons of the SCI regenerative models of the Correia PD et al bioinformatics study? As mentioned above, *RPRM* genes are mainly expressed in the central nervous system of vertebrates, including in the spinal cord[46]. We hypothesize that RPRM may be down-regulated in neurons to improve neuronal survival, and aim to further test its effects on neuronal differentiation.

2. Goals

A previous analysis of three transcriptomics studies of spinal cord injury in rodents subjected to different therapies, found 68 differentially expressed genes (DEGs) in common. The main goal of this work was to develop molecular tools to study the role of two of these potentially neuroregenerative genes, *MPP3* and *RPRM*, found up-regulated and down-regulated, respectively, at an early (acute) post-lesion period. Particularly interesting it will be to understand why is *RPRM*, a gene associated to cell cycle inhibition, down-regulated in neuroregenerative conditions in neuronal non-mitotic cells.

The specific aims of this study were to:

- Clone the *MPP3* full length gene in a mammalian expression vector, to be able to overexpress it in neuroblastoma cells, for future proliferation and neuritogenic assays;
- Implement the CRISPR/Cas9 technique in the lab, by designing and testing molecular tools for genome editing of neuroblastoma cells, first targeting *RPRM*, a gene potentially associated with neuroregeneration;
- Establish a stable line of neuroblastoma cells expressing SpCas9, that can be used in the future to modulate various potentially neuroregenerative genes via CRISPR/Cas9 techniques.
- Expand our dataset of potential neuroregenerative genes by performing a comparative bioinformatics analysis of a previous analyzed study and a new study with a subacute transcriptomic dataset of the lesion site tissue of a SCI animal model with improved spinal cord regeneration and functional recovery.

3. Materials and Methods

3.1. *MPP3* experimental study

3.1.1. The *MPP3* plasmids

Two vectors carrying *MPP3* cDNAs – 'pGEM-T MPP3 flag full form' (carrying the *MPP3* full length, for prokaryotic expression) and 'pBABE CMV NEO MPP3' (with a truncated *MPP3* CDS; for mammalian expression) were kindly provided by Prof. Dr. Jan Wijnholds (Dept. of Ophthalmology, Leiden University Medical Center, The Netherlands). These plasmids carry an ampicillin resistance gene selection marker.

These cDNAs, sent on sterile filter paper, were firstly amplified for stock maintenance, and purified. To recover the approximately 6 µg of each plasmid, the DNA-carrying filters spots were cut, and each placed in a 1.5 mL tube, to which 100 µL of TE buffer was added (Invitrogen). The tubes were briefly vortexed, incubated at room temperature (RT) for 5 min, further briefly vortexed and centrifuged for a few seconds. After that, 5 µL of the supernatants were used to transform *Escherichia coli* (*E. coli*) bacterial cells as below indicated. The reminder of the filter paper/TE extract was stored at -20ºC.

3.1.2. Heat shock transformation of competent bacteria

Competent *E. coli* DH5α cells (NZYTech) were thawed on ice and 5 µL of the plasmid DNA to amplify were added to 30 µL of competent cells. Cells were incubated on ice for 30 min, posteriorly subjected to heat shock by incubation in a 42ºC bath for 90 seconds, and then rapidly transferred to ice for 3 min, allowing the DNA to be incorporated. 970 µl of LB medium was added to the cells, followed by incubation at 37ºC 200 x g for 60 min, for bacteria to recuperate. Cells were finally centrifuged at 1700 x g for 1 min, the supernatant discarded, and the pellet resuspended in the remaining volume to be plated in LB/agar plates (Luria-Bertani (LB) Broth, NZYtech; Agar Granulated, NZYtech), complemented with 50 µL/mL of ampicillin (NZYTech) with the help of glass beads. Plates were incubated at 37ºC overnight. Colonies were picked and inoculated in 5 mL of LB medium with ampicillin for 16 hours at 37°C with orbital shaking at 180 x g (starter culture). This cell suspension was then grown in 200 mL of LB medium with ampicillin overnight under the same conditions, for plasmid amplification and extraction following the NucleoBond® Xtra Midi (Macherey-Nagel) protocol. The protocol was performed under a sterile environment and the appropriate controls were used.

3.1.3. Midiprep with the NucleoBond® Xtra Midi for DNA extraction

Plasmids were extracted from 200 ml bacterial cultures, which were inoculated with a single bacterial clone in LB medium with the appropriate antibiotic (ampicilin for *MPP3* plasmids), and grown overnight at 37ºC and 180 x g. When the optical density reached approximately 0.2,

bacterial cells were pelleted by centrifugation at 6000 x g for 15 min at 4ºC, and supernatants were discarded. Bacterial cells were completely resuspended, lysed, neutralized and centrifuged. Plasmids were then purified following the NucleoBond® Xtra Midi (Macherey-Nagel) protocol and the concentration of extracted DNA samples was analyzed with Nanodrop (DeNovix).

3.1.4. Strategy for subcloning the *MPP3* full-length cDNA insert

Plasmid 'pGEM-T MPP3 flag full form' (hereon simplified to pGEM-T fl-MPP3), of prokaryotic expression, encodes the full-length MPP3 protein (Figure 3.1.A.). In contrast, the eukaryotic expression pBABE CMV NEO MPP3 (hereon simplified for pBABE t-MPP3), encodes an unstable protein, truncated after the SH3 domain (MPP3ΔGuK) (Figure 3.1.B.). Thus, to be able to modulate this gene in neuroblastoma cells, it was necessary to plan a subcloning procedure to insert the full-length *MPP3* gene into the eukaryotic expression vector.

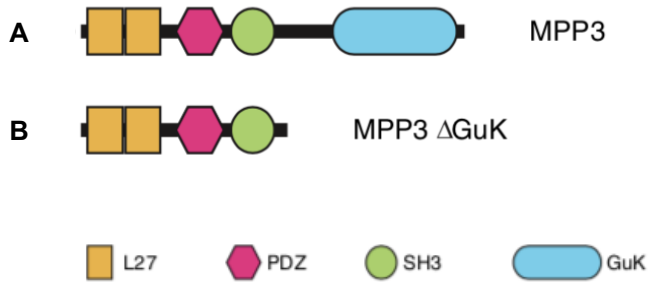


Figure 3.1. Schematic protein structures of MPP3 (A) and MPP3ΔGuK (B) homologues. As a member of the membrane palmitoylated protein family, MPP3 contains two L27 domains, one PDZ, SH3 and GuK domain. Adapted from Kantardzhieva A[41].

After analyzing the genomic maps of the plasmids, it was realized that both could be digested with *Bam*HI and *Sal*I (both from Thermo Fisher Scientific), cutting at one third of the *MPP3* CDS (*Bam*HI) and after the end of the *MPP3* CDS (*Sal*I) (Figure 3.2.).

To make the connection of interest, both plasmids ('donor' plasmid pGEM-T fl-MPP3 and 'receptor' pBABE t-MPP3 plasmid), were digested with these enzymes. The first released the insert of interest, and the second released an unwanted fragment with one third of the truncated and the whole *GUK* CDS. The remaining larger fragment of the pBABE t-MPP3 plasmid would be the receptor vector. The product of this subcloning technique was named 'pBABE CMV NEO MPP3 full' (circular / 8077 bp) (Figure 3.2.).

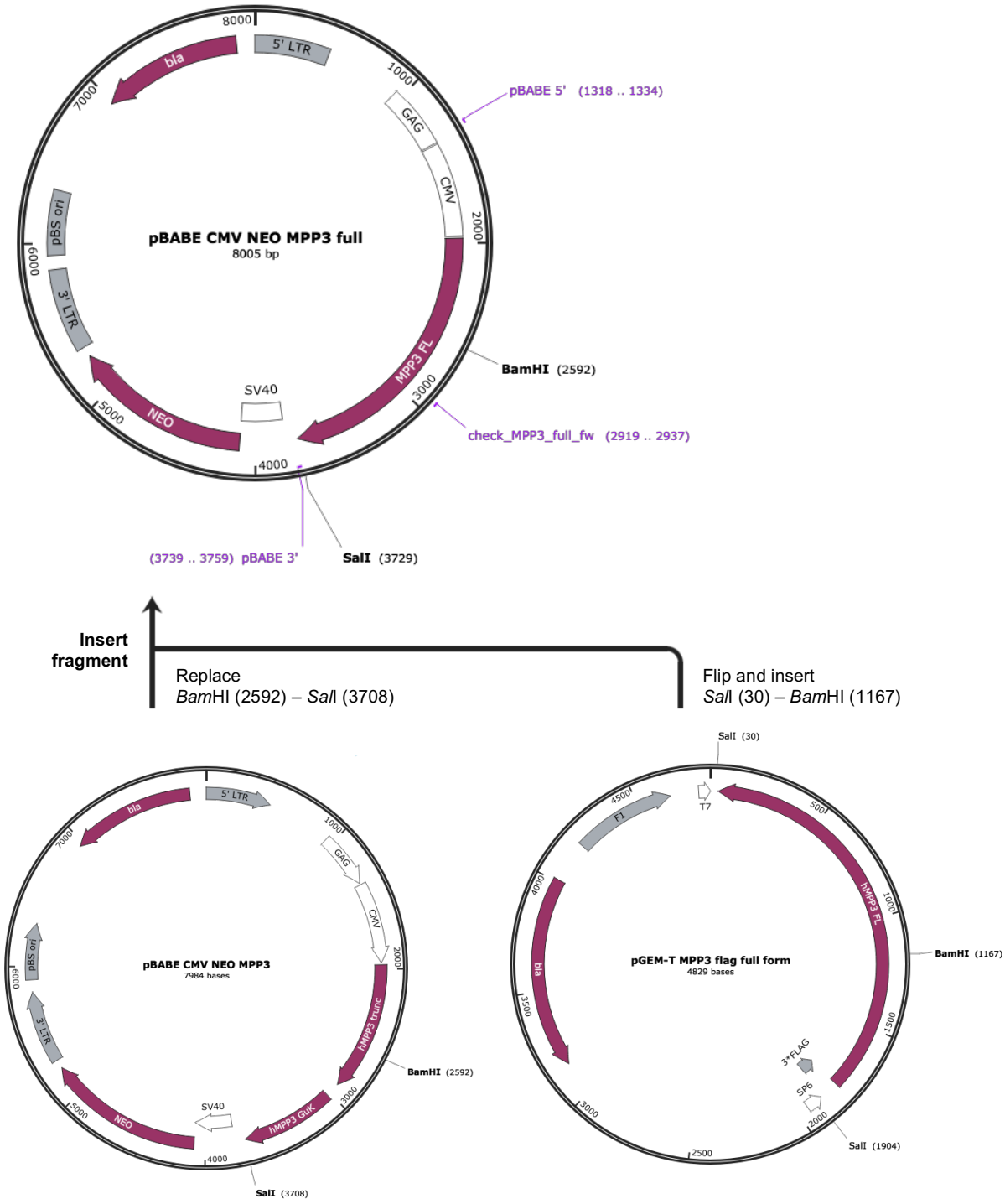


Figure 3.2. Scheme of the subcloning strategy to create the recombinant pBABE CMV NEO MPP3 full, from the plasmids pBABE CMV NEO MPP3 and pGEM-T MPP3 flag full form. Scheme constructed with the aid of the SnapGene 5.2. software.

3.1.5. Plasmids enzymatic digestion

For the enzymatic digestion of each of the *MPP3* plasmids, a Master Mix was prepared, as indicated in table 3.1., for a final volume of 20 µL and using 1 µg of DNA. Digestion occurred at 37ºC, for 7 hours.

Table 3.1. Components of the enzymatic restriction reactions for *MPP3* subcloning.

Reagent	pGEM-T (for insert)	pBABE (for receptor)
DNA (1ug)	10 µL	10 µL
BamHI buffer 10x	2 µL	2 µL
*Bam*HI	1 µL	1 µL
*Sal*I	2 µL	2 µL
dH_2O	5 µL	5 µL

3.1.6. Agarose gel electrophoresis and gel extraction

Digested samples were loaded and separated in 1% agarose (GRS Agarose LE, GRiSP) gel in 1x TAE buffer (Tris-Acetate-EDTA, Fisher BioReagents) at 100 V for 40 min. Using GeneRuler DNA Ladder Mix (Thermo Fisher Scientific), with double-stranded DNA in the range of 100 bp to 10,000 bp on agarose gels, the bands of interest (~7 kb from the digested pBABE and ~1,1 kb from the digested pGEM-T) were rapidly excised from the gel under the UV light with a scalpel, followed by isolation and purification with NucleoSpin® Gel and PCR clean-up (Macherey-Nagel), as indicated in the manufacturer's protocol. The binding buffer added to each sample was left overnight, at RT, to increase the yield. At the end, DNAs were eluted with 50 µL of the Elution Buffer, previously heated to 50ºC.

3.1.7. Ligation and amplification of the *MPP3* recombinant DNA

After the purification of both DNAs, the concentration was measured and followed by the ligation of the fragments. The calculation for the ligation was made with the help of the online tool 'Ligation Calculator' (http://www.insilico.uni-duesseldorf.de/Lig_Input.html), choosing a 1:3 insert:vector ratio. The ligation was prepared with 5 µL of vector (9.5 ng/µL), 5 µL of insert (6 ng/µL), 5 µL dH_2O and 5 µL of T4 DNA ligase Anza (Invitrogen), for 1 hour at RT. 5 µl of the product of this reaction was later used to transform competent bacteria, as described in section 3.1.2. Two transformations were made.

3.1.8. *MPP3* primers selection for positive colonies and successful cloning confirmations

MPP3 primers were designed for confirmation of positive colonies (to differentiate between bacteria transformed with re-ligated pBABE plasmid from the ones transformed with the

recombinant pBABE plasmid). The recombinant plasmid would have, in theory, only 21 bp more than the initial pBABE plasmid. Thus, to distinguish between them and confirm positive colonies by colony PCR, a forward primer (named 'check_MPP3_full_fw') was designed for that 21 bp sequence and used in conjunction with the common pBABE 3' reverse primer. A PCR amplicon of 840 bp was expected when using these two primers and the recombinant pBABE DNA as template. For the subsequential sequencing of the recombinant plasmid, the pBABE 5' forward primer was also chosen, to pair with the pBABE 3' reverse one. For primer design for PCR and DNA sequencing, a number of rules were followed. Primers should have at least 15-17 nucleotides and avoid the formation of the secondary "hairpin" structure, have a percentage of cytosine and guanine nucleotides greater than 50% and an annealing temperature between 50-65ºC. The chosen primers can be found below, and were acquired at Eurofins Genomics.

***MPP3* primers designed/used in this study** (from Eurofins Genomics)**:**

Forward 'check_MPP3_full_fw': 5' GATCAGCCTTGTGACAAAG 3' (19 bp)	Tm(ºC)=54.5
Reverse 'pBABE 3'_rv': 5' ACCCTAACTGACACACATTCC 3' (21 bp)	Tm(ºC)=57.9
Forward 'pBABE 5'_fw': 5´CTTTATCCAGCCCTCAC 3' (17 bp)	Tm(ºC)=52.8

3.1.9. Colony-Polymerase Chain Reaction (PCR) screening

From each plate with transformants with putative recombinant pBABE CMV NEO MPP3 full plasmid, 20 colonies were analyzed, in a total of 42 colony-PCR reactions (40 with potential positive colonies and 2 negative controls). For this purpose, a Master Mix solution was prepared for approximately 48 reactions (please see **Table 3.2.**). This Master Mix is a ready-to-use mixture of DNA polymerase, salts, magnesium, and dNTPs for efficient PCR amplification. The PlatinumTM SuperFi^{TM} (Invitrogen) proofreading DNA polymerase has the PlatinumTM hot-start technology, which inhibits DNA polymerase activity at ambient temperatures, allowing RT reaction setup and storage of pre-assembled PCR reactions for up to 24 hours prior to the PCR. Enzyme activity is restored after the initial denaturation step.

Table 3.2. Components of the colony-PCR reaction for *MPP3* positive colonies screening.

Reagent	Volume (µL):
Platinum™ Superfi™ PCR Master Mix 2X (polymerase + salts + dNTP's)	300
Primer forward	30
Primer reverse	30
dH_2O	240
Total volume	**600**

Once prepared, 12.5 µL of this Master Mix was added to each of the 42 PCR microtubes. Then, in sterile conditions, one colony was picked at a time, transferred to a new LB-Agar-Ampicillin plate (to allow the future identification and growth of the positive colonies) and to the respective PCR tube. Finally, the samples were placed in a thermocycler with the corresponding definitions for the Platinum™ Superfi™ polymerase and for the 840 bp we intended to amplify, which are identified in blue in the Table 3.3. After the PCR, gel loading buffer [DNA Gel Loading Dye (6X), Thermo Fisher Scientific] was added to each PCR product, mixed, and the samples briefly centrifuged, and further analyzed using agarose gel electrophoresis.

Table 3.3. Colony-PCR conditions, for MPP3 positive colonies screening.

Step		**Temperature:**	**Duration:**
Initial denaturation		98ºC	30''
30 PCR cycles	Denaturation	98ºC	10''
	Anneal	52ºC	10''
	Extend	72ºC	30''
Final extension		98ºC	5'

3.2. Mammalian cells culture and manipulation

3.2.1. SH-SY5Y and 293T cells culture and maintenance

SH-SY5Y (ATCC® CRL-2266™) is a twice subcloned cell line derived from the SH-SY subclone of the parental SK-N-SH cell line, isolated from a human neuroblastoma[52,53]. This cell line contains two phenotypes: neuroblast-like cells ("N"-type), and epithelial-like cells ("S"-type)[52]. This cell line is popular as a model for neuronal differentiation, since cells can be converted to a functional neuronal-like phenotype (characterized by neuronal markers)[52,54].

Human embryonic kidney (HEK) 293T cells (ATCC® CRL-3216™) were generated from HEK293 cells, stably transfected with a plasmid expressing a temperature-sensitive version of the SV40 large T antigen[55,56]. The HEK293 cell line itself was originally derived from human embryonic kidney cells transfected with mechanically sheared DNA of human adenovirus type 5 (Ad5)[55,56].

Neuroblastoma SH-SY5Y cells were cultured in Minimum Essential Medium (MEM):F12 (1:1) containing 10% Fetal Bovine Serum (FBS), 2 mM L-glutamine (Gibco) and 1% antibiotic/antimycotic mix (Thermo Fisher Scientific). Kidney 293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) high glucose (4.5 g/L) supplemented with 10% FBS and 1% (100U/mL) of antibiotic/antimycotic mix. Cells were maintained in culture at 37°C in a humidified atmosphere containing 5% CO_2. Cells were cryopreserved at -80ºC in 20% FBS and 5%(v/v) Dimethyl Sulfoxide (DMSO) Hybri-Max™ (Sigma Aldrich), at a cell density of 80-90% of confluency per vial.

Cells were cultured and split for maintenance by changing the medium two to three times per week, when confluence achieved approximately 90%. To do so, cells were washed with Phosphate Buffer Saline (PBS) 1X (BupH Modified Dulbecco's PBS Packs, Thermo Fisher Scientific) and further incubated with 1 mL of Trypsin-EDTA (0.05%, 1X), phenol red (Gibco) for about 3 to 5 min in the humidified incubator to detach cells. When individual cells separated and detached from the dish surface, cells were resuspended in 4 mL of culture medium, which deactivates trypsin. This suspension was centrifuged at 1,000 rpm for 3 min at RT, and the supernatant was discarded. The pellet of cells was then resuspended in 10 mL of fresh medium. If cells were only to be maintained, usually 1 mL of this new cell suspension was plated in 100 mm plates with 9 mL of fresh medium in a 1:10 ratio. If we were about to start an experiment, cells were scored in a hemocytometer, using the vital Trypan blue dye (Sigma-Aldrich), to plate an exact number of cells.

3.2.2. Mammalian cells count

Cell number was calculated by counting cells in a hemocytometer (BRAND® counting chamber BLAUBRAND® Neubauer improved, Sigma-Aldrich) under 100X magnification in a brightfield microscope. Cell concentration per mL was calculated by the expression: ~x * dilution factor * $1x10^4$ = cells/mL (~x, the average number of cells counted in the 4 squares of the hemocytometer; $1x10^4$, the hemocytometer volume factor to mL). Cell viability was assessed by counting cells from an aliquot of cell suspension stained with trypan blue dye. The non-colored cells were regarded as live cells, while blue ones were compromised or dead cells.

3.2.3. Test and compare the efficiency of different transfection reagents

To test which of the three available transfecting reagents could be more efficient in the transfection of *MPP3* cDNA, we compared the efficiency of the transfection reagents Turbofect Reagent (Thermo Fisher Scientific), Lipofectamine 2000 (Thermo Fisher Scientific), and X-treme Gene 360 (Sigma-Aldrich) with 1 or 2 µg of DNA (V1_EGFP[57]). $1x10^5$ SH-SY5Y cells were plated 24 hours before transfection and, on the day of transfection, the medium was aspirated and replaced with 2 mL of fresh medium (complete MEM). Two wells were also used as negative controls, where transfection was not performed. DNAs and reagents were diluted separately and joined later to form the complexes, in a 1:3 ratio, respectively. Right before their use, reagents were shortly vortexed. As recommended by the fabricants, 1 µg of DNA was diluted in 100 µL of optiDMEM Reduced Serum Medium (1X) (Thermo Fisher Scientific) or with 2 µg of DNA in 200 µL. All mixtures were gently vortexed before incubation. The ones with Turbofect or X-treme Gene were incubated for 20 min, while the ones with Lipofectamine 2000 were incubated for only 5 min. After incubation, the mixtures were equally and gently distributed in the respective wells. 48 hours after transfection, cells were fixated using 1 mL of 4% Paraformaldehyde solution (PFA) (Sigma Aldrich) for 15 min followed by three washes with PBS and then the slides were prepared using Vectashield Antifade Mounting Medium with DAPI (Vector Laboratories). Epifluorescence microscopy analysis of the slides was carried out using an OlympusIX81 microscope.

3.2.4. *MPP3* transient transfection of SH-SY5Y cells by X-treme Gene 340 reagent

In order to study the role of MPP3 in neuritogenesis, the expression of the recombinant *MPP3* cDNA was first tested in neuroblastoma cells, following transient transfection using X-treme Gene 360. SH-SY5Y cells were plated in six-well plates until reaching ~75% confluency. 2 µg of DNA were diluted in 200 µL of serum-free growth medium. After briefly vortexed, 6 µg of X-treme Gene were added to the 2 µg of DNA. The mixture was vortexed and incubated at RT for 15-20 min. During this period, cell medium was replaced by fresh medium in the 6-well plate. The mixture was subsequently added drop by drop to each well, with gentle agitation to achieve even distribution. Cells were incubated for 24h and 48h at 37º in a CO_2 incubation. As negative control of overexpression of *MPP3*, empty pCMV vector was used.

3.2.5. Immunocytochemistry assays and microscopy analysis

In order to qualitatively analyze the over-expression of *MPP3* in neuroblastoma cells, an immunocytochemistry protocol was performed. Firstly, cells grown on coverslips and transfected with *MPP3* cDNA as above indicated were fixated using PFA at 4% for 15 min, followed by three washes of 5 min each with PBS. Cells were subsequently permeabilized and blocked with 3% Albumine Bovine Fraction V (BSA) (NZYTech) and 0.3% Triton (Calbiochem, Merck) in PBS solution, for 1 hour. Following, MPP3 was labeled by incubating cells for 2 hours with a rabbit anti-MPP3 antibody (14650-1-AP, ThermoFisher Scientific), diluted 1:60, together with a mouse anti-beta III Tubulin antibody (T8578, Sigma), diluted 1:500, to label the microtubules, in PBS with 3% BSA and 0.3% Triton. Cells were further washed 3 times with PBS, for 5 min each, and incubated with Alexa488-conjugated anti-mouse antibody (Invitrogen) and Alexa594-conjugated anti-rabbit antibody (Invitrogen), both diluted 1:300 in PBS with 3% BSA and 0.3% Triton. Cells were further washed 3 times with PBS, for 5 min each, and 1 last time with Mili-q water. Negative controls were cells incubated with blocking solution and secondary antibodies only (no primary antibodies), to check non-specific binding. The slides were prepared using Vectashield Antifade Mounting Medium with DAPI, and epifluorescence microscopy analysis of MPP3 was carried out using a Zeiss Axio Imager Z1.

3.3. CRISPR/Cas9 experimental study

3.3.1. Selection of the CRISPR/Cas9 system

To proceed with the knockout of *Reprimo*, we opted to use two lentiviral vectors to assemble the CRISPR system, the pLentiCas9-Blast (Genscript), which contains the sequence coding for the SpCas9 nuclease, and the pLentiGuide-Puro (Genscript), which contains the guideRNA (gRNA) that should identify the *RPRM* gene in the cell genome. pLentiCas9-Blast was firstly used to create a stable SH-SY5Y cell line expressing SpCas9, to be latter transduced with the virus containing the pLentiGuide-Puro, to target the interest gene to knockout.

PLentiCas9-Blast (Figure 3.3. A) is a 12.8 kb lentiviral vector, prepared to be assembled with a second or third generation lentiviral system. In addition to the elements mentioned above, this vector also has bacterial replication origin and ampicillin resistance gene to select positively transformed bacteria. In addition to this resistance, it also has genes for resistance to blasticidin and bleomycin for selection of transduced cells.

PLentiGuide-Puro (Figure 3.3. B), with 8.3 kb, has, like pLentiCas9-Blast, the LTR sequences, the elements for replication and selection in bacteria, and also resistance to puromycin, for selection of positively transduced cells.

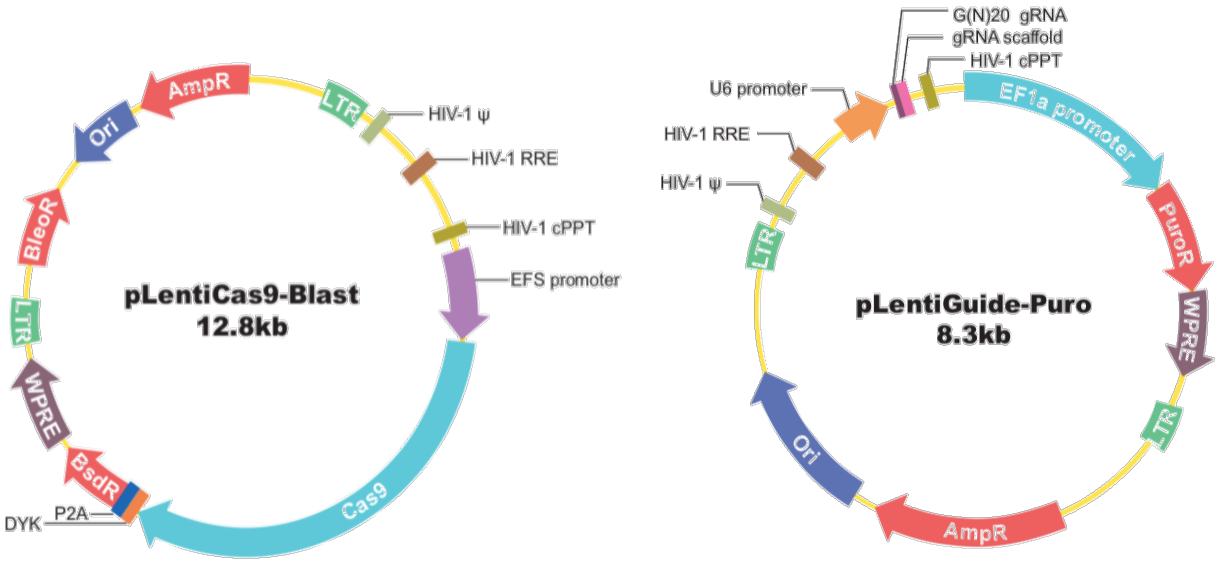


Figure 3.3. pLentiCas9-Blast (A) and pLentiGuide-Puro (B) vectors' maps. Taken from Genscript.

Regarding pLentiGuide-Puro, Genscripts' presented 6 gRNA options for *RPRM* gene, developed and validated by Feng Zhang's laboratory at the Broad Institute (please see Table 3.4.).

Since no more information about these gRNAs was available, the Deskgen software was used to characterize these guides. This platform has a variety of useful tools for designing and analyzing guides for CRISPR technology, namely, the Guide Picker. Using this tool, *RPRM* was chosen as the gene of interest, and *Homo sapiens* (38.81) as the genome in which the editing would be executed. As an output, this tool gave some of the most important scoring parameters and functions about the potential gRNAs. Next, the gRNA options from Genscript were search in that output, to collect the information about each one, which can be found in the table below:

Table 3.4. Characteristics of available gRNAs for human *RPRM* knockout.

Sequence	PAM	Representation	Percent peptide	Off-target (HSU 2013)	On-target (Doenche 2016 positionless)
~~**gRNA 1**~~ TAGCCTGTACATAATGCGCG	TGG	1/1	49	100	62
~~***gRNA 2***~~ CAGGCCCGCCACGTCCGTC	TGG	1/1	8	99	45
~~***gRNA 3***~~ GCACCACGCGCATTATGTAC	AGG	1/1	47	100	57
~~**gRNA 4**~~ TCCCTCCGCGAAGCCGTCGT	CGG	1/1	35	100	57
gRNA 5 GGTGACCGACGACGGCTTCG	CGG	1/1	37	99	63
~~**gRNA 6**~~ CAATCTGCTCATCAAGTCCG	AGG	1/1	76	97	58

Off-target score represents the specificity of the gRNA, with a higher score meaning fewer predicted sites in the genome where this gRNA is likely to cut. The on-target activity score also predicts the ability of the gRNA to cut DNA at the intended target site using data from several large-scale CRISPR experiments and more complex bioinformatic tools that analyze several desirable/undesirable sequence traits. Percent peptide refers to the guide position within the protein coding portion of the entire gene. The base pair values for this sequence are normalized from 0-100, 5'–3' to provide percentage progression through the CDS. gRNA activity does not correlate strongly to percent peptide, but to targeting consensus exons, functional domains and avoiding the extreme 5' and 3' ends of the gene, which have been shown to enhance gRNA activity ([52,53]). Having all these in mind, we pursued a guide with a percentage of on-target as high as possible, with greater than 55% being considered a good percentage; an off-target percentage close to 100; and a guide that directs SpCas9 to a sequence relatively close to the start of the CDS, in order to decrease the probability of resulting in an expressed peptide with some functionality.

When observing the characteristics of each guide (**Table 3.4.**), gRNAs 1, 3 and 6 were excluded for having higher percent peptide, and gRNA 2 for having an on-target less than 55%. From here, it was a matter of balance: although gRNA 4 has a slightly lower percentage of peptide than gRNA 5, and an off-target of 100, the latter, in addition to having percentage of peptide and off-target very similar to gRNA 4, had the highest percentage of on-target of these two options - 63, and was thus chosen as the *RPRM* guide present in the pLentiGuide-Puro, acquired from Genscript.

3.3.2. Amplification of pLenti DNAs

pLentiCas9-Blast and pLentiGuide-Puro, sent in the form of 4 µg of lyophilized plasmid DNA, were prepared as indicated in their datasheets, for "Standard delivery". Firstly, the tubes containing the plasmids were centrifuged at 6,000 x g for 1 min at 4ºC. Next, 20 µL of sterilized water were added at each plasmid, and vortexed for 1 more min. There was no need to heat the solution at 50ºC for 15 min to dissolve the DNA. Vials were stored at -20ºC until used for bacterial transformation. These vectors were amplified using the same protocol that was used for *MPP3* plasmid DNA (section 3.1.2.), using competent *Escherichia coli* DH5α cells.

3.3.3. Establishment of a stable neuroblastoma cell line expressing SpCas9

Before proceeding with the *RPRM* knockout itself, other main objective of this work was to establish a stable neuroblastoma cell line expressing SpCas9 (SH-SY5Y-Cas9), to be used in future work of the group, and high-content screen assays using CRISPR/Cas9 as genetic editing tool. Figure 3.4. and Figure 3.5. summarize the procedures and timeline, respectively, to obtain stable SH-SY5Y-Cas9 cells.

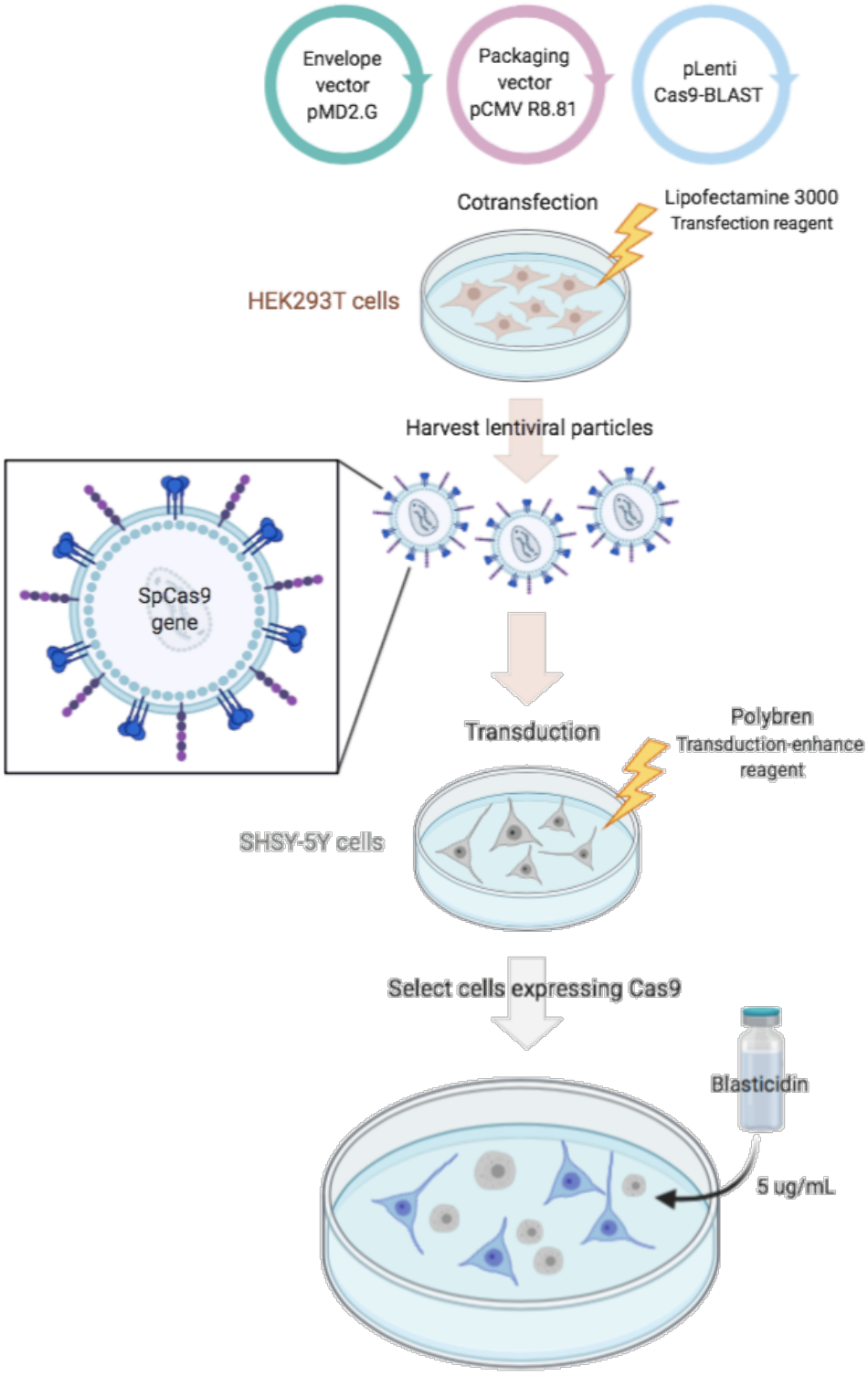


Figure 3.4. Production of *SpCas9*-carrying lentivirus like particles (LVLPs) for neuroblastoma cells transduction. LVLPs were first produced by HEK293T cells, co-transfected with a packaging plasmid (pCMV R8.81), an envelope plasmid (pMD2.G), and a transfer vector (pLentiCas9-BLAST), using Lipofectamine 3000 as transfection reagent. Cells conditioned medium containing LVLPs was harvested to transduce SH-SY5Y cells, with the help of the transduction-enhancing reagent, Polybrene. Transduced cells were select via blasticidin resistance. The schemes were created with BioRender.com.

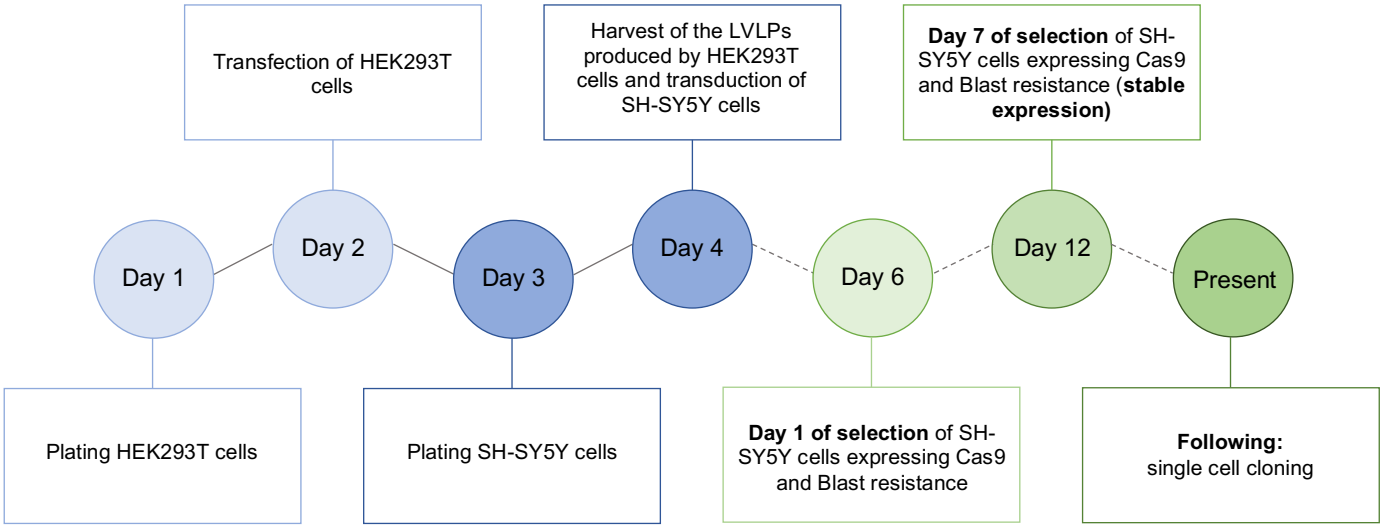


Figure 3.5. Timeline of the experiments regarding the establishment of a stable neuroblastoma cell line expressing SpCas9.

3.3.4. Polybrene toxicity test and optimization of blasticidin and puromycin concentrations for selection of the transduced SH-SY5Y cells

Polybrene is cationic polymer that enhances the efficiency of the retroviral/lentiviral infection to the mammalian cells, neutralizing the charge repulsion between virions and the cell surface (https://www.sigmaaldrich.com). However, this transduction reagent can be toxic to cells when at higher concentrations. Therefore, before the transduction of SH-SY5Y with the lentivirus containing the pLentiCas9-BLAST DNA, it was necessary to ensure that the concentration of polybrene (Santa Cruz Biotechnology) we were planning to use to help transduction would not kill cells *per se*. To do so, 8 µg/mL of the polymer were added to SH-SY5Y cells grown in 2 mL of MEM, in a well of a 6-well plate.

The same concern was placed regarding antibiotics cytotoxicity, and the working concentrations of blasticidin and puromycin to be used to select the positively transduced cells, were also optimized. Three concentrations were tested for each antibiotic. After a literature search, the concentrations chosen to test were 2.5 µg/mL, 5 µg/mL and 10 µg/mL for blasticidin[58,59] (Blasticidin S hydrochloride; Alfa Aesar) and 1 µg/mL, 2 µg/mL and 4 µg/mL, for puromycin[60,61]. Neuroblastoma cells were plated in 6-well plates ($2x10^5$ cells/well), 24 hours before adding the antibiotics.

3.3.5. Production of *SpCas9*-gene containing lentiviral particles

To produce *SpCas9* containing lentiviral particles, HEK293T cells were co-transfected with the three lentiviral plasmid vectors (packaging, envelope and transfer, with the last one being pLentiCas9-BLAST). The packaging and envelope plasmid vectors, pCMV R8.81 and pMD2.G, respectively, were kindly provided by Prof. Dr. Daniela Ribeiro (iBiMED).

The HEK293T cell line was chosen for the lentivirus production because it expresses the Large T antigen that is essential for replicating plasmids containing an SV40 origin of replication at a high copy number, in the transfected cell. Additionally, these cells show increased cell growth and are easily transfected, supporting high-level of expression of viral proteins[56].

Since HEK293T cells are easily lifted, a coating procedure with poly-L-ornithine was performed on day 1, before their seeding. Firstly, 10 mL of the poly-L-ornithine (Sigma Aldrich) diluted in sterile water solution was added to the plate, distributed uniformly, and incubated at 37ºC. After 2 hours, the excess solution was aspirated, and the plate rinsed 2 times with sterilized water and 1 time with DMEM. HEK293T cells ($3x10^6$ cells) were then seeded in the coated 10 cm plate.

Transfection was performed 24 hours later, on day 2. LipofectamineTM 3000 Reagent (Lipofectamine 3000 Transfection Reagent, Life Technologies) was diluted in 400 µL of Optim-MEM and vortexed. The positive surface charge of the liposomes mediates the interaction of the adsorbed nucleic acids and the cell membrane, overcoming its electrostatic repulsion. The DNA mix was prepared by diluting 8 µg of plasmid DNA (1 µg of pCMV R8.81, 3 µg of pMD2.G and 4 µg of pLentiCas9-BLAST) in 400 µl Opti-MEM with P3000TM Reagent and mixing well. This mix was then added to the solution containing LipofectamineTM 3000 Reagent, in a 1:1 ratio.

The volume of Opti-MEM used (400 µL for each mix, 800 µL in total) was based on a 50:1 ratio of Opti-MEM (µL) to total DNA (µg). The volume of P3000TM Reagent used (16 µL) was based on a 2:1 ratio of P3000TM Reagent (µL) to total DNA (µg), and the volume of LipofectamineTM 3000 Reagent (12.03 µL) used was based on a 1.33:1 ratio of LipofectamineTM 3000 Reagent (µL) to P3000TM Reagent (µL).

After homogenized, the final mix was incubated for 15 min at RT and added to the cells in a dropwise manner. During the incubation time of the transfection mixture, HEK293T media was carefully substituted by DMEM with 10% FBS and without antibiotic-antimycotic solution.

On day 3 medium was changed to complete DMEM medium. In the same day, SH-SY5Y cells were seeded on a 6-well plate, at a density of $2.4x10^5$ cells/well, with 3 wells per condition (3 wells were used for transduction and the other 3 as negative controls). On day 4, the LVLP-containing HEK293T conditioned media was filtered through a 0.45 µm filter.

3.3.6. Transduction of SH-SY5Y cells with the *SpCas9* gene-containing lentivirus

Still on day 4, to perform the transduction of SH-SY5Y, 8 µg/mL of polybrene were added to 9 mL of the filtered medium and, after removing the medium of the SH-SY5Y cells plated the day before, 3 mL of the prepared medium were carefully added to each one of the three wells aimed for transduction. In the three wells aimed for negative controls of the transduction, 3 mL of SH-SY5Y new complete medium, containing 8 µg/mL of polybrene, were distributed for each well.

24h after transduction, on day 5, the medium of each well was replaced with new SH-SY5Y complete medium.

3.3.7. SH-SY5Y-Cas9 transduced cells selection (ongoing)

On day 6, SH-SY5Y growth medium was replaced with selective medium containing blasticidin, at a final concentration of 5 µg/ml to start the selection of cells positively transduced, until all the negative control cells died. On days 8, 9 and 10, the medium was replaced with new medium containing 5 µg/ml of the antibiotic, to remove dead cells and toxic debris. On day 11, the medium was also replaced but, this time, cells were previously washed with PBS.

3.4. Comparative bioinformatics analysis of transcriptomic SCI studies

3.4.1. Selection of transcriptomic studies

With an initial aim of updating the original bioinformatics study of Correia PD et al[37], literature searches for transcriptomic studies reporting improved spinal cord regeneration and functional recovery after SCI, in rodent models, after a given treatment, were carried out on the dataset databases Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/), Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra/), and Array Express (https://www.ebi.ac.uk/arrayexpress/), and on the literature database Pubmed (https://pubmed.ncbi.nlm.nih.gov/), between June and September of 2020. The Pubmed database was used to guaranty that all studies with expression data and obeying to the parameters set, could be found even if their expression data was not uploaded into the three transcription databases described.

A combination of the following keywords was used: “spinal cord injury therapy” OR “spinal cord regeneration” OR “spinal cord injury transcriptomics”, with filters for *Mus musculus* species or *Rattus* genus.

Besides the three studies already analyzed by Correia PD et al[37], only one extra study, obeying to our set parameters, was found (Zeng H et al, 2019[40] PMC6854783), that analyzed the transcriptome of the lesioned area at 28 days post-injury, dpi. Its findings were thus compared to one of the three original studies, E-GEOD-69334[38], that performed transcriptomic analyses at similar time points (20 and 30 dpi).

3.4.2. Data extraction

The available processed expression data files of the two selected studies were extracted. The microarray expression profile E-GEOD-69334[38] based on Affymetrix GeneChip Rat Genome 230 2.0 array (Santa Clara, California, EUA) platform was obtained from Array Express database from EMBL. Regarding the Zeng H et al, 2019[40] study (PMC6854783), there was only access to the RNA-sequencing raw data (https://submit.ncbi.nlm.nih.gov/subs/sra/SUB5917896/overview) deposited in the SRA, or to the DEGs lists made by the authors, based on the following parameters: p-adjust< 0.05 and $|\log2FC| \geq 1$.

3.4.3. Identification of Differential Expressed Genes (DEGs)

Since the new PMC6854783 study included for analysis had already established parameters for the selection of DEGs from the transcriptomic data of the lesion site at 28 days dpi – significant adjusted p-value and $|log2FC| \geq 1$ – the same stringent parameters were applied to the extracted 20 dpi datasets of the E-GEOD-69334 study, for comparison ends.

DEGs from E-GEOD-69334 were identified by first calculating the fold-change (FC) difference between injured animals that received the regenerative treatment and animals only injured (that received no intervention, here used as controls), and then filtering for the thresholds described above (p-adjust< 0.05 and $|log2FC| \geq 1$). Up or down-regulated genes were separated in two distinct groups for each study. Reactome (https://reactome.org) was the online tool used to obtain the FC, and results were further filtered using Microsoft Excel.

Reactome is an open-source database that provides bioinformatics tools for an easier visualization and analysis of biological pathways' information (https://reactome.org).

ReactomeGSA (Version 73, released on June 17, 2020) can be accessed through Reactome's web interface, and supports five types of quantitative 'omics data: microarray intensities, transcriptomics raw and normalized read counts, and proteomics spectral counts and intensity-based quantitative data. These different types of data are processed using two different methods: statistics for discrete quantitative data (in case of raw transcriptomics read counts and spectral counting based quantitative proteomics data) and statistics for continuous data (https://reactome.org).

Discrete values are normalized using edgeR's calcNormFactors function, and the data is transformed using limma's voom function. Continuous data is directly processed using limma and normalized using limma's normalizeBetweenArrays function.

The study analyzed in this platform, E-GEOD-69334, was based on Affymetrix GeneChip Rat Genome 230 2.0 array (Santa Clara, California, EUA) platform and was obtained from Array Express database from EMBL. To proceed with the analysis, firstly, Reactome was accessed. Among the three different algorithms to use for the differential pathway analysis, "Camera" was selected to perform a differential expression analysis between two groups of samples, with "Microarray normalized" data. To be uploaded, data files must be a tab-delimited text file (CSV or TSV file), where the first column contains the gene or protein identifiers and the first row contains the sample names.

The data files of E-GEOD-69334 for analysis consisted in one CSV file containing 6 datasets, with 3 being from the lesion site of injured and NT3 treated animals at the 20 dpi, and 3 from lesion control animals (that did not receive the treatment) at the same timepoint. Once the dataset was added, it was proceeded to the annotation of the experimental metadata, in order to define the groups for the differential expression analysis that followed. Since the goal was to compare gene expression between these conditions, the data from each NT3 treated sample were

annotated as “Treatment” and data from the lesion control animals were annotated as “Control”. Then, “Condition” was selected as the comparison factor and set “Control” and “Treatment” as “1st group” and “2nd group”, respectively. ReactomeGSA then creates a Microsoft Excel report of the results, including the Identifier, logFC, AveExpr, t, P.Value, adj.P.Val, and B.

Finally, to convert the I.D. references from the microarray expression profiles from the study E-GEOD-69334 to the respective gene names, The Synergyzer web database (http://llama.mshri.on.ca/synergizer/translate) was used. The ID list was uploaded, and the filters were chosen accordingly: ensemble (2015-11-17) for Authority, *Rattus norvegicus* for Species, affy_rat230_3 for “FROM” namespace, and rgd_symbol for “TO” namespace. IDs for which the corresponding gene names were not found, were removed from the list for further analysis.

3.4.4. Finding Human orthologs

Normalization of the DEGs to the respective human orthologs was made by using BioMart from Ensemble (release 101) (www.ensembl.org/biomart/martview/), a web-based tool referred before, that allows normalization of data with different identification codes as well as the finding of homologue or ortholog genes. As filters input, the ID list with the gene names was uploaded, and human orthologs were selected as desired output. Only genes with human homology type “one2one” or “one2many” and human orthology confidence of 1 were further included in this study. Additionally, the online tool jVenn (jvenn.toulouse.inra.fr/) was used to remove the duplicates of the filtered orthologs.

DRSC Integrative Ortholog Prediction Tool (DIOPT) (https://www.flyrnai.org) (Version 8.0 (Aug 2019)) was another tool we tested to find the human orthologs.

Comparing BioMart with DIOPT, it was possible to notice DIOPT’s outputs contained more duplicates and less orthologs, therefore, BioMart was chosen to find the human orthologues.

3.4.5. DEGs comparisons, Gene Ontology, pathways enrichment and protein-protein interaction network analyses

A Venn diagram, using the Jvenn software (jvenn.toulouse.inra.fr/), was created to perform comparisons between the up- and down-regulated DEGs from the two studies, to search for common DEGs.

To create networks of protein-protein interactions (PPI) and to find potential enriched pathways and GO terms within it, the online Search Tool for the Retrieval of Interacting Genes/Proteins database (STRING 11.0, https://string-db.org) was used. The found up-regulated genes common to the 2 datasets were submitted to analysis, and String's PPI network parameters were established so the networks would be based on Textmining, Experiments, Databases, Co-expression, Neighborhood, Gene fusion and Co-occurrence data, with connection interaction scores of at least 0.4 and without expansion of the network.

To get more knowledge about each one of the 12 up-regulated genes common to the two studies, two other databases were used: The Universal Protein Resource (UniProt, https://www.uniprot.org), to collect data from three categories of Gene Ontology: biological process, molecular function and cellular component, and the Database for Annotation, Visualization and Integrated Discovery (DAVID 6.8., https://david.ncifcrf.gov/tools.jsp), to retrieve information about the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome Pathways, associated to each DEG.

Additionally, a literature search was done to collect more information about the function of each gene.

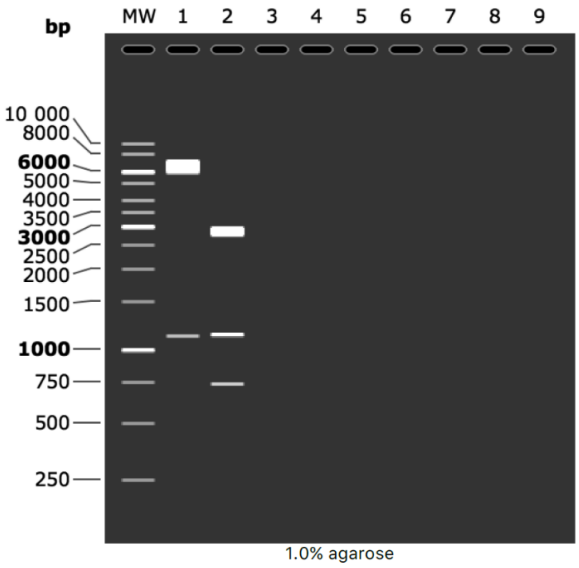
4. Results

4.1. MPP3 overexpression experimental studies

This section describes the results from the experimental studies on MPP3, particularly the subcloning of the full-length gene from a prokaryotic expression vector to a eukaryotic expression vector, the monitorization of positive colonies by 'colony PCR', and the transfection of the recombinant plasmid into neuroblastoma SH-SY5Y cells.

4.1.1. Construction of a *MPP3* full-length mammalian expression vector

Plasmids pGEM-T MPP3 flag full form ('pGEM-T fl-MPP3'), for prokaryotic expression of the MPP3 full length protein, and pBABE CMV NEO MPP3 ('pBABE t-MPP3'), for eucaryotic expression of a truncated MPP3 protein, were kindly provided by Prof. Dr. Jan Wijnholds. To be able to 'reconstruct' a *MPP3* full-length sequence to be expressed in eukaryotic cells, both plasmids were digested with the restriction enzymes *Bam*HI and *Sal*I. A scheme of a 1.0% agarose electrophoretic gel with the expected bands from the digestions is presented (Figure 4.1.), as well as the real gel obtained (Figure 4.2.). For the pBABE t-MPP3 plasmid, 2 bands were expected and obtained, one with of 6940 bp [*Sal*I (3780) − *Bam*HI (2664)] and another of 1116 bp [*Bam*HI (2664) − *SalI* (3780)]. The first contained the majority of the vector backbone, and was the 'receptor vector', while the smaller one contained part of the *MPP3* gene. For the pGEMT fl-MPP3 vector, 3 bands were expected and obtained, one of 2955 bp [*Sal*I (1904) − *Sal*I (30)], other of 1137 bp [*Sal*I (30) − *Bam*HI (1167)], and one of 737 bp [*Bam*HI (1167) − *Sal*I (1904)]. The 1137 bp band contained the partial sequence of the non-truncated *MPP3* gene that would serve as the 'insert for ligation'.



1. Lane 1: pBabe CMV NEO MPP3 *Bam*HI + *Sal*I
    a. 6868 bp = *Sal*I (3708) − *Bam*HI (2592)
    b. 1116 bp = *Bam*HI (2592) − *Sal*I (3708)
2. Lane 2: pGEM-T MPP3 FLAG full form *Bam*HI + *Sal*I
    a. 2955 bp = *Sal*I (1904) − *Sal*I (30)
    b. 1137 bp = *Sal*I (30) − *Bam*HI (1167)
    c. 737 bp = *Bam*HI (1167) − *Sal*I (1904)

Figure 4.1. Scheme of expected results for the enzymatic restriction of the MPP3 plasmids. Expected band profile for a 1.0% agarose gel with the digestion reaction of pBabe t-MPP3 (lane 1) and pGEM-T fl-MPP3 FLAG (lane 2) with *Bam*HI and *Sal*I. Predicted band profile made with the SnapGene 5.2. software.

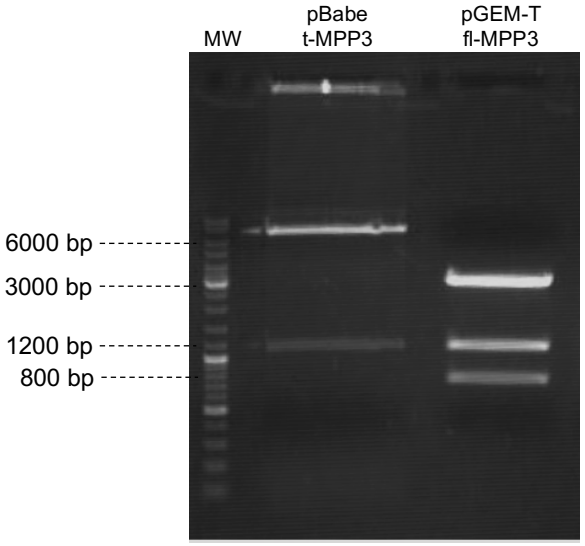


Figure 4.2. Results obtained for the enzymatic restriction of the MPP3 plasmids. DNA bands profile obtained in a 1.0% agarose gel electrophoresis after digestion of pBabe t-MPP3 (lane 1) and pGEM-T fl-MPP3 (lane 2) with BamHI and SalI. The gel run at 100 V for 40 minutes in 1x TAE running buffer.

The bands of interest (~7 kbp from the digested pBABE and ~1,1 kbp from the digested pGEM-T) were excised from the gel, isolated and purified using the NZYGelpure kit (NZYTech). The ligation was performed for 1 hour at RT, at a 1:3 proportion. The final product of this ligation is an 8035 bp vector of eukaryotic expression, named by us pBABE CMV NEO MPP3 full (hereon simplified to pBABE fl-MPP3, containing the coding sequence (CDS) for the expression of a full length MPP3 protein.

The product of this ligation was amplified by transformation in competent bacteria (*Escherichia coli* DH5α cells (NZYTech) and selected with Ampicillin for 24h. Various positive colonies were obtained in the MPP3 cDNA transformed cells. Several of these were monitored by colony-PCR, using the primer forward 'check_MPP3_full_fw', which only hybridizes with the recombinant MPP3 sequence, and the primer reverse 'pBABE 3'_rv'. The expected product of the amplification with these primers has 840 bps and spans the region of the *MPP3* CDS that was now recombined as in the correct full length *MPP3*. Figure 4.3. presents the electrophoresis agarose gels obtained.

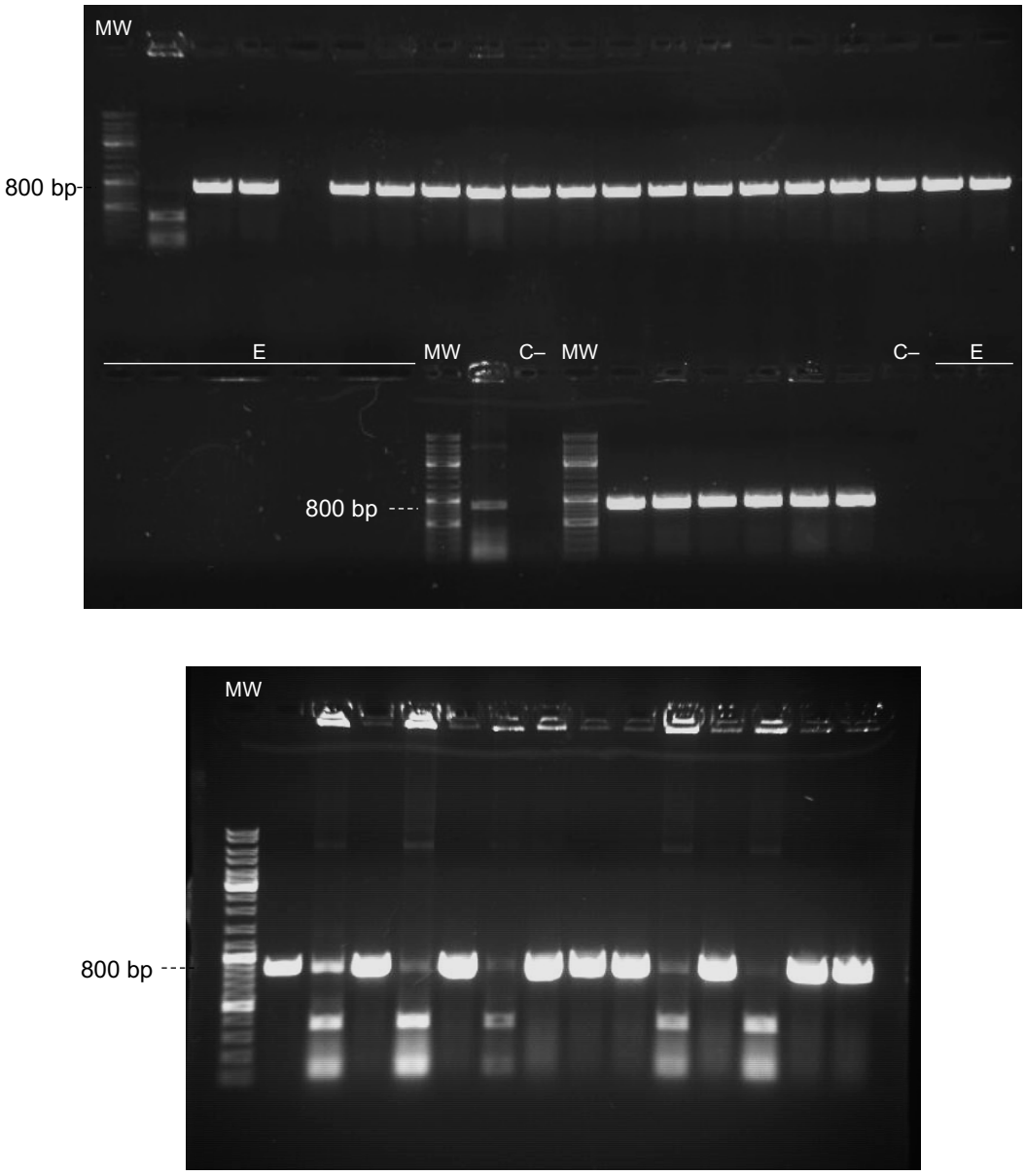


Figure 4.3. Gel electrophoresis with the results from the colony-PCR screening for pBABE fl-MPP3 transformants. 1.0% agarose gel loaded with the colony-PCR products, showing that most of the positive colonies tested contained the expected pBABE fl-MPP3 ligation product, as they express the 840 bp amplicon expected for the designed primers. The lanes identified as 'MW' ('molecular weight') correspond to the ones loaded with GeneRuler DNA Ladder Mix (Thermo Fisher Scientific); the lanes identified as 'C' correspond to the ones loaded with the samples with no DNA, and the lanes identified as 'E' (empty) correspond to wells that were not loaded. The lanes with more than one band correspond to samples that did not contain the recombinant MPP3 cDNA.

As it can be seen in Figure 4.3., and as expected, all the lanes loaded with the negative control samples (with no colony/DNA added to the PCR mixture), identified as "C", are clear. Additionally, 32 out of the 40 lanes that were loaded with samples from the colony-PCR reactions containing genetic material, have amplified the sequence of interest (840 bps), which only exists in the recombined vector pBABE fl-MPP3 full (and not in the originally digested pBABE t-MPP3 vector). The other 15 samples probably contained genetic material from colonies that did not incorporated the recombination vector but the re-ligated vector.

4.1.2. DNA sequencing of the pBABE fl-MPP3 cDNA

Figure 4.4. shows the alignment between the sequencing readout of the selected positive clone DNA, and the pBABE fl-MPP3 and the pBABE t-MPP3 sequences. The sequencing readout completely aligned with the sequence expected to be in the pBABE fl-MPP3, confirming that a full length *MPP3* construct was successfully obtained.

The full sequencing with the pBABE 5' and 3' primers, which covered around 2/3 of the *MPP3* CDS, can be found in Annex (Supplementary Figure 1.). Within the *MPP3* sequence, no mismatch was observed between the sequencing readout and the known *MPP3* CDS.

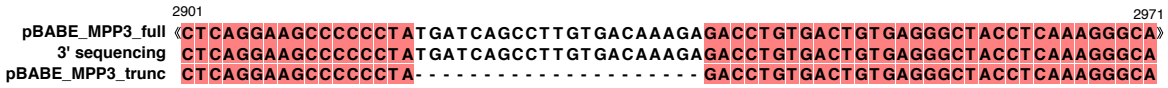


Figure 4.4. Confirmation of the pBABE fl-MPP3 DNA sequence. Alignment between the sequencing readout of the selected positive clone DNA (pBABE_MPP3_full) and the pBABE fl-MPP3 (3' sequencing) and pBABE t-MPP3 (pBABE_MPP3_trunc) sequences.

4.1.3. Optimization of neuroblastoma cells transfection

Before proceeding with the transfection of SH-SY5Y cells with the newly produced recombinant pBABE fl-MPP3 cDNA, the efficiency of different transfection reagents – Turbofect Reagent, Lipofectamine 2000, and X-tremeGene 360 was first tested and compared. Cells were transfected with different DNA concentrations of the EGFP vector for 48 hours, as schematized in Figure 4.5. V1_EGFP was the DNA chosen for this assay because it contains a reporter gene, the (*enhanced) Green Fluorescent Protein* (*EGFP*), that encodes for a protein that emits green fluorescence when exposed to the light in the blue to ultraviolet range. This facilitates the analysis of the transfection efficiency, since only the successful transfected cells will emit fluorescence.

	1. Turbofect	2. Lipofectamine 2000	3. X-tremeGene 360
A 1 µg DNA	A 1 µg V1_EGFP 3 µl Turbofect	C 1 µg V1_EGFP 3 ml Lipo	E 1 µg V1_EGFP 3 ml X-treme
B 2 µg DNA	B 2 µg V1_EGFP 6 µl Turbofect	D 2 µg V1_EGFP 6 µl Lipo	F 2 µg V1_EGFP 6 µl X-treme

Figure 4.5. Scheme of the transfection conditions tested in a 6-well plate. Three different transfection reagents: Turbofect, Lipofectamine 2000 and X-tremeGene360.

Following 48 hours of transfection, the cells were monitored by epifluorescence microscopy (Figure 4.6). In the microphotographs, from various fields of view, it is possible to observe and compare the proportion of transfected cells (in green), relatively to the total amount of cells (transfected and not transfected (in blue, stained with the nuclear marker DAPI)). The

microphotographs from the transfections with Lipofectamine show a lower total number of cells than those in which the other transfection reagents were tested, although the green marking is intense. This may indicate some type of toxicity for this transfection reagent. On the other hand, as mentioned before, both Turbofect and XtremeGene do not appear to have affected cell survival. However, comparing these two reagents, it is possible to verify a greater number of transfected cells using X-tremeGene. Considering these observations, it was decided to proceed with the transfection of 2 µg of pBABE fl-MPP3 with X-TremeGene.

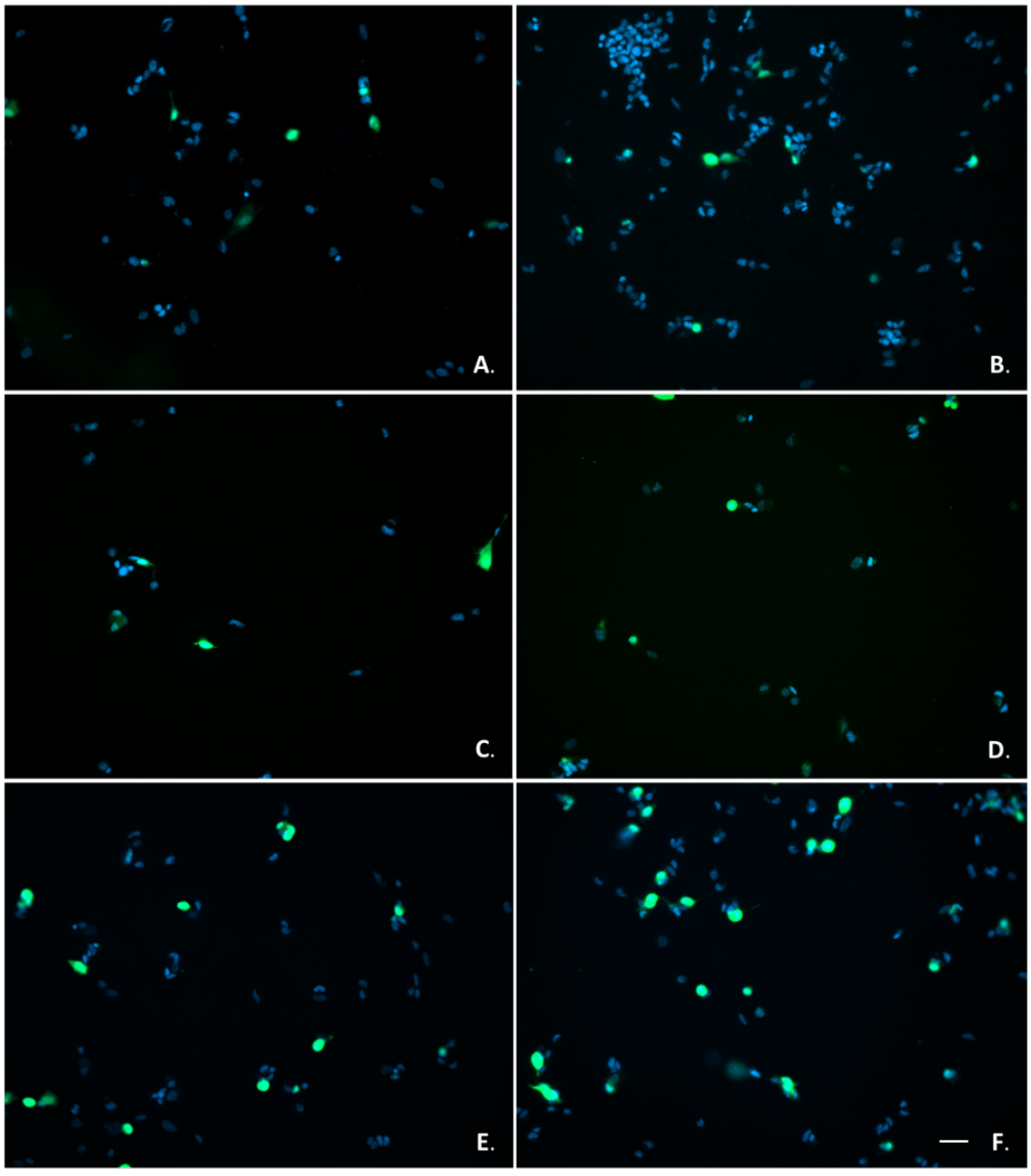


Figure 4.6. Epifluorescence microphotographs for qualitative comparison of the efficiency of three different transfection reagents. Microphotographs A. and B. correspond to Turbofect; C. and D. to Lipofectamine 2000; and E. and F. to X-tremeGene 360. Microphotographs A, C, and E correspond to cells transfection with 1 µg of V1_EGFP, while in B, D, and E, cells were transfected with 2 µg of the same plasmid. DAPI, in blue, is marking the nucleus. Green fluorescence results from the expression of the reporter gene, the green fluorescent protein (GFP), confirming the successful transfection of the vector. Bar = 50 µm.

4.1.4. MPP3 transient transfection of SH-SY5Y cells by X-treme Gene 340 reagent

The recombinant *MPP3* cDNA was transfected in neuroblastoma cells, using X-treme Gene 360, in a DNA to reagent ratio of 1:3. Following 48 hours of transfection, the cells were monitored by epifluorescence microscopy (Figure 4.7).

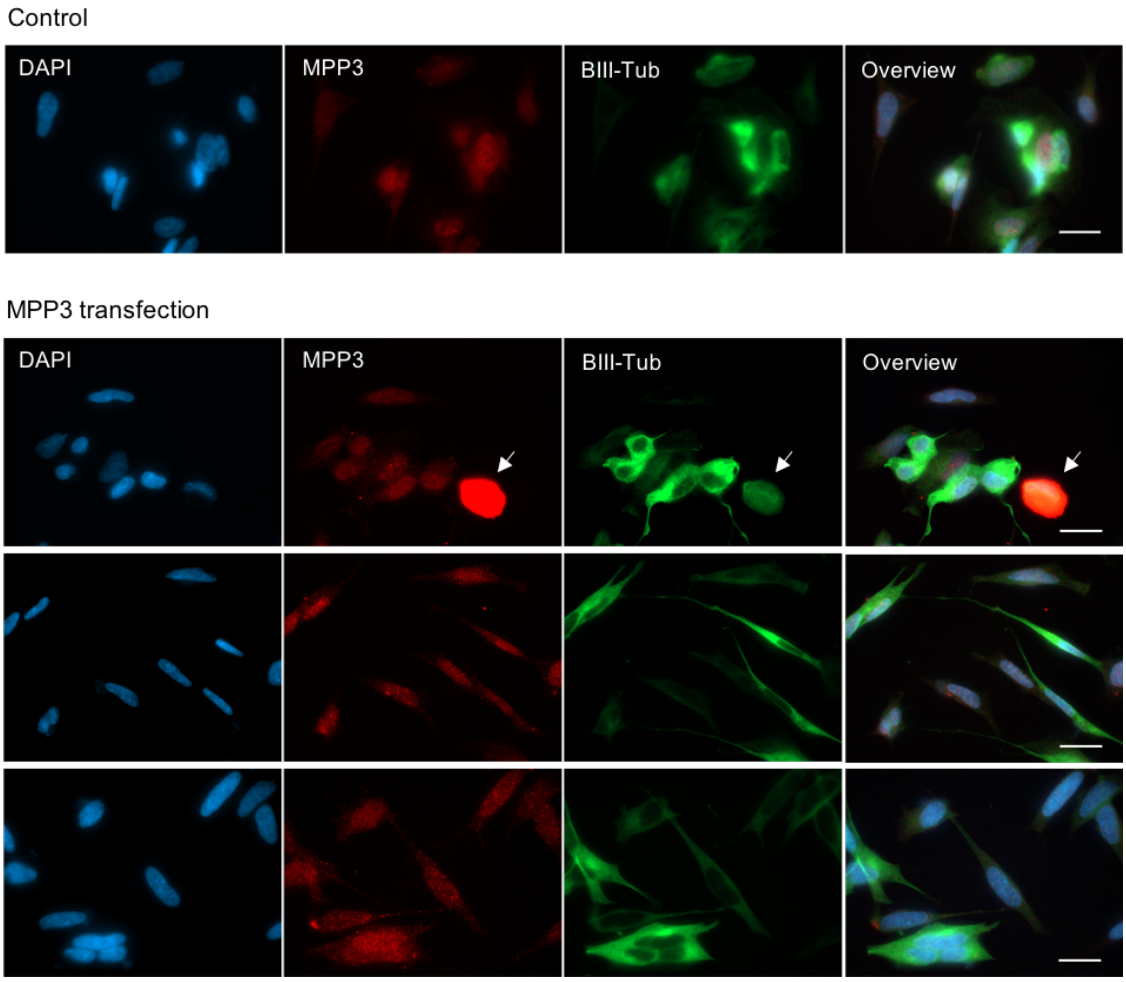


Figure 4.7. Epifluorescence microphotographs of the transient transfection of SH-SY5Y cells with pBABE fl-MPP3. Upper panel: negative control cells, transiently transfected with an empty pCMV vector MPP3. Lower panel: cells, transiently transfected with the pBABE fl-MPP3 plasmid. The X-treme Gene 340 reagent was used. BetaIII-tubulin (in green) stains neuronal microtubules. DAPI (in blue) is marking the nucleus. Arrows indicate transfected cell overexpressing MPP3. Bar = 20 µm.

DAPI (blue) marks the cell nuclei, Anti-mouse Alexa Fluor 488 (green fluorescence), detects the mouse anti-beta III Tubulin antibody, which stains neuronal microtubules, and anti-rabbit Alexa Fluor 594 (red fluorescence), detects rabbit anti-MPP3 antibody, labeling MPP3.

Only on the microphotographs corresponding to the transfection of *MPP3* condition (Figure 4.7.), a reduced number of cells, already dead, with a very intense red fluorescence marking, was observed (identified in Figure 4.7. by the arrows).

Moreover, it seems there was no correlation between the staining of MPP3 and the staining of beta-III-Tubulin, both in transfected and non-transfected cells. It was also analyzed whether

MPP3 was more expressed in cells with a more differentiated phenotype (neuronal-like) (two lower panels from Figure 4.7.), to search for a potential relationship between MPP3 expression and differentiation, but this was also not demonstrated.

4.2. *RPRM* knockout-related experimental studies

This subsection describes the results from the experimental procedures whose final aim was to create a stable neuroblastoma SH-SY5Y cell line expressing SpCas9, for subsequent gene knockout assays by transduction with different gRNA-containing viral particles, such as the RPRM gRNA lentivirus.

4.2.1. Polybrene, Blasticidin and Puromycin cytotoxicity tests for SH-SY5Y cells

Before the transduction of SH-SY5Y cells with the lentivirus containing the pLentiCas9-BLAST DNA, it was necessary to ensure that the concentration of polybrene planned to be used would not kill cells per se. To do so, 8 µg/mL of the polymer were added to SH-SY5Y cells grown in 2 mL of DMEM, in a well of a 6-well plate. Cells were visually monitored for a few days, and cells grown normally (as parallel control cells), and never showed signs of suffering cytotoxic effects. Thus, the 8 µg/mL concentration of polybrene did not represent an insult for the cells, and was used as reported by other authors[62,63].

Since the pLentiCas9-BLAST and pLentiGuide-Puro vectors contain blasticidin and puromycin resistance genes, respectively, those antibiotics would be used for future selection of successful transductions. Blasticidin was used for monitoring successful transduced SH-SY5Y-Cas9 cells and puromycin would be used for successful transduced SH-SY5Y-Cas9-gRPRM cells. Hence, it was necessary to first perform antibiotic dose-response assays to choose the best concentration, since too low antibiotic dose may not be enough to kill the cells and too high concentration could lead to too fast cell toxicity, killing even successfully transduced populations. Based on literature and on the manufacturers' instructions, three concentrations were tested for each antibiotic: 2.5, 5 and 10 µg/mL of blasticidin; and 1, 2 and 4 µg/mL of puromycin. To monitor the "normal" cell growth, negative controls (no antibiotic added) were also monitored.

Cells' response to blasticidin (Figure 4.8.) was monitored until day 5, when essentially all the cells exposed to 10 µg/mL blasticidin were dead or already at an apparent apoptotic stage (microphotographs in Figure 4.8). Based on this experiment, the 5.0 µg/mL blasticidin concentration was chosen, given that it was effective in selecting a high percentage of cells upon 5 days, similarly to 10 µg/mL blasticidin, and had already been suggested by other authors[58,64,65].

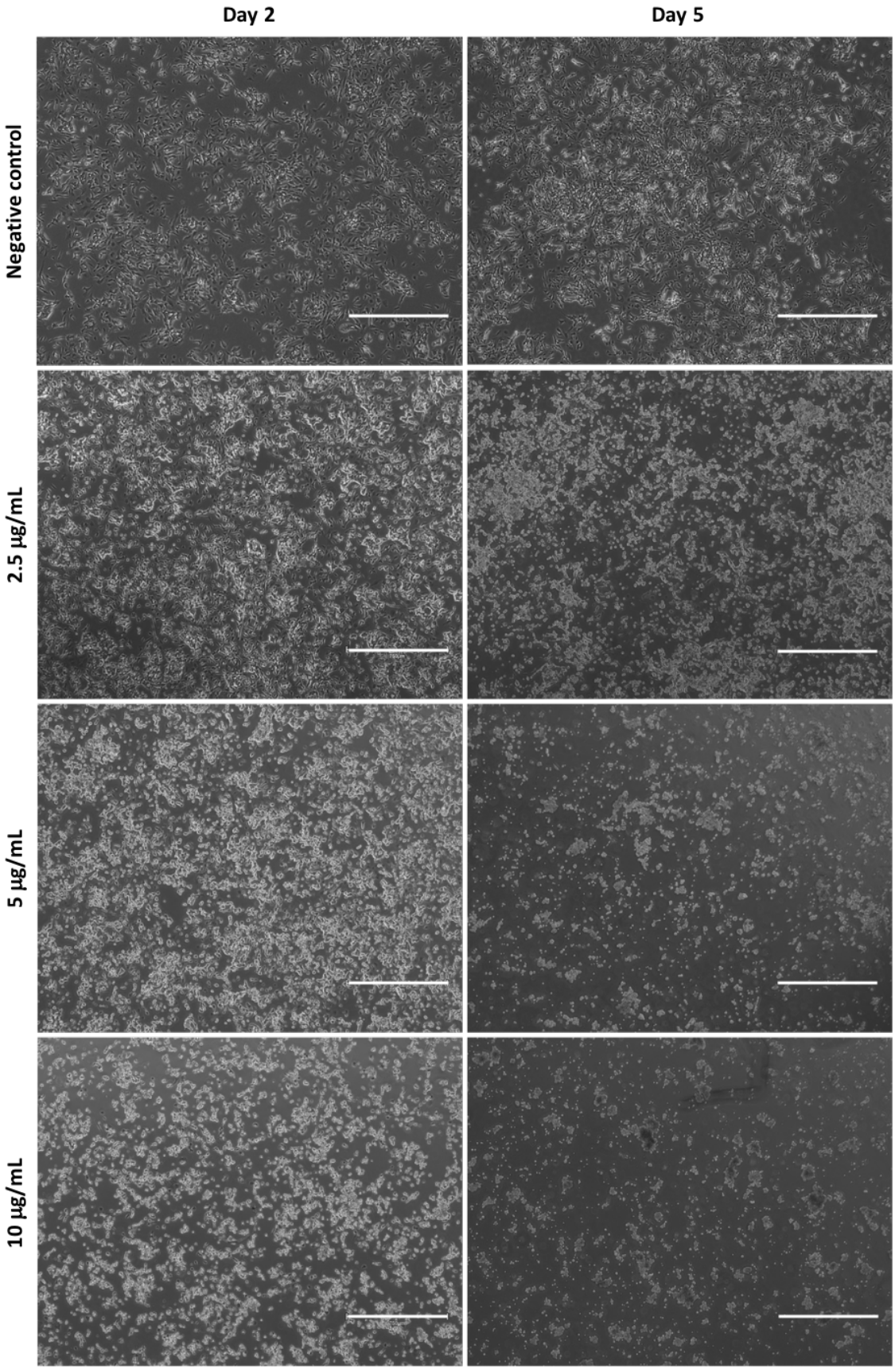


Figure 4.8. Phase contrast microphotographs of a blasticidin dose-response experiment in SH-SY5Y cells, exposed for up to 5 days to 2.5, 5.0 and 10 µg/mL of blasticidin. Microphotographs from control cells came from another parallel experiment. Bar = 750 µm.

Regarding puromycin (Figure 4.9.), all the cells in the 2 or 4 µg/mL conditions died after 48h of exposure. Cells exposed to 1 µg/mL died more slowly, but this concentration showed a good efficacy. This is in accordance with other authors using puromycin in SH-SY5Y cultures[66,67]. The 1 µg/mL puromycin concentration was chosen to later select the cells integrating the gRPRM.

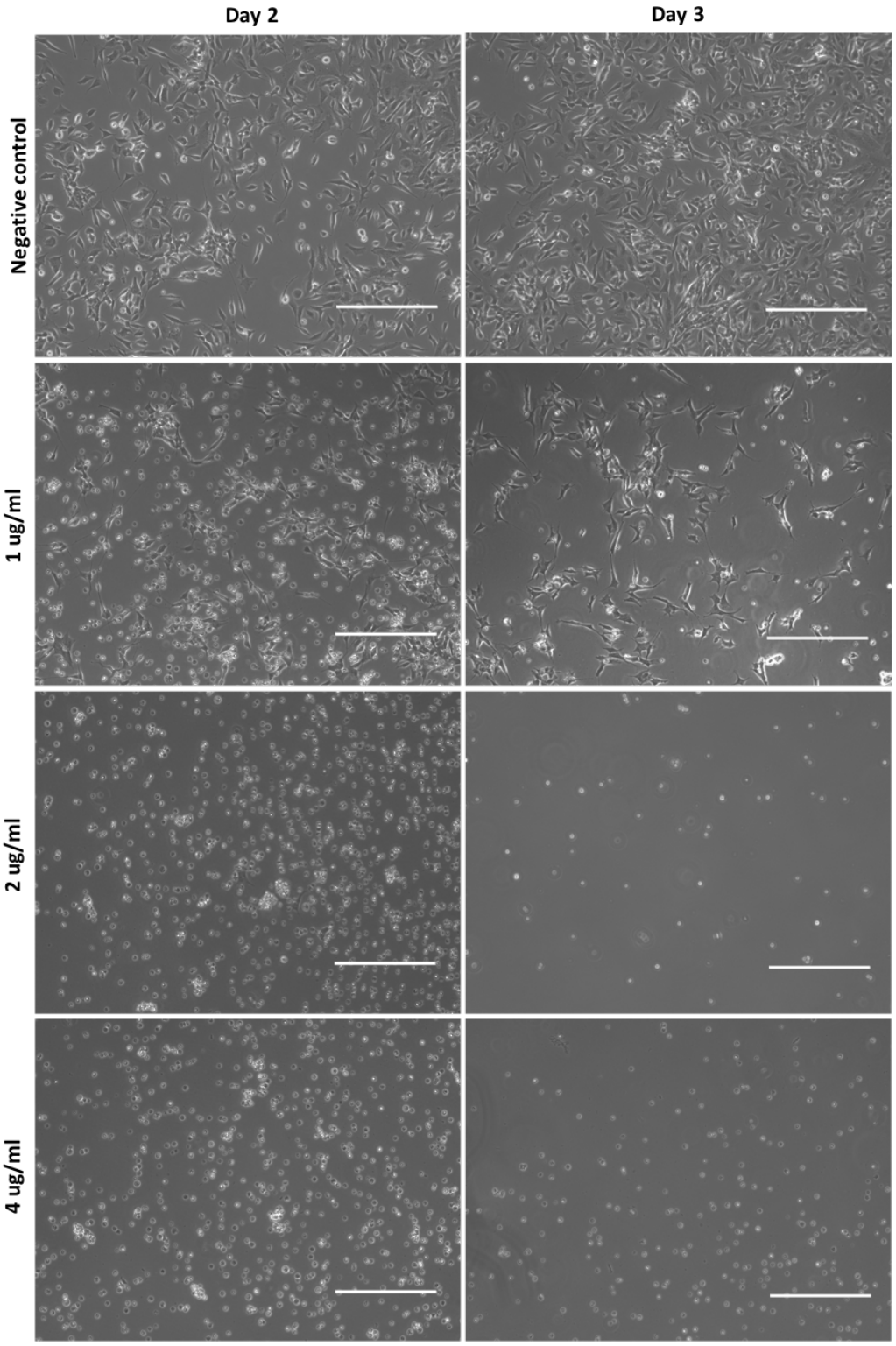


Figure 4.9. Phase contrast microphotographs of a puromycin dose-response experiment in SH-SY5Y cells, exposed for up to 5 days to 1, 2 and 4 µg/mL of puromycin. Negative control - cells not exposed to the antibiotic. Bar = 300 µm.

4.2.2. Selection of SH-SY5Y cells transduced with the *SpCas9*-containing lentivirus

In order to establish a neuroblastoma cell line stably expressing the SpCas9 protein, to be further used as a cellular tool for the knockout of RPRM and of other potentially neuroregenerative genes, *SpCas9*-containing lentiviral particles were first produced. For that, HEK293T cells were co-transfected with the three lentiviral plasmid vectors (packaging, envelope and transfer, with the last one being pLentiCas9-BLAST). After 2 days, the LVLPs produced by HEK293T cells were harvested and used to transduce SH-SY5Y cells.

2 days after transduction, the culture medium was replaced with selective medium containing 5 µg/ml of blasticidin to start selecting cells positively transduced with the *SpCas9*-containing lentivirus. Assay was performed until all negative control cells died, on the 7th day of selection (Figure 4.10.).

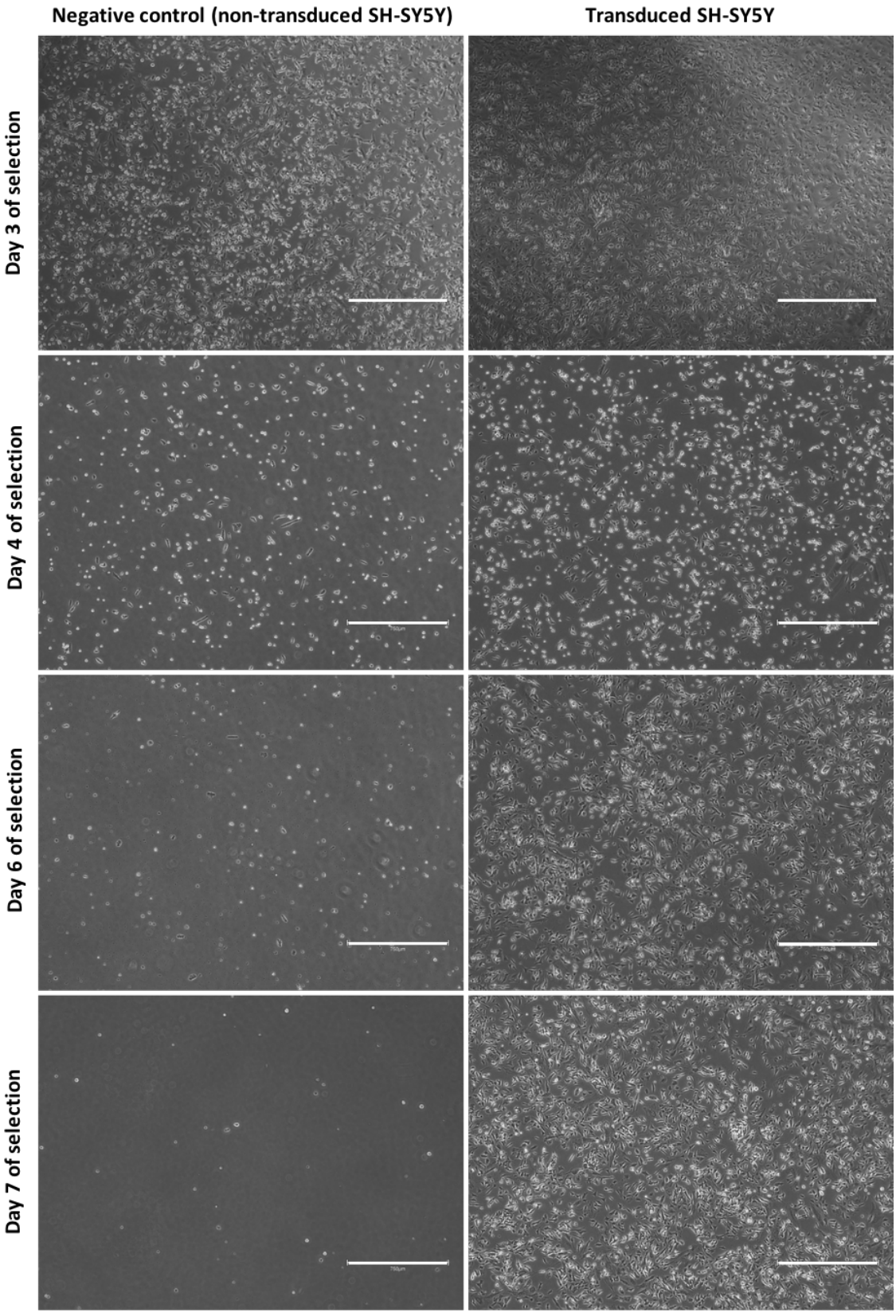


Figure 4.10. Selection of positively transduced SH-SY5Y-Cas9 neuroblastoma cells with 5 µg/mL of blasticidin for 7 days. At day 7 of selection, all the non-transduced cells were dead. Bar = 750 µm.

On day 3 of selection, it was already possible to notice a higher level of cell death on the negative control plate, in which transduction was not made. Although it was also possible to observe apoptotic (or already dead) cells in the transduced cells plate, it was noticeable a greater number of well adhered SH-SY5Y cells, both N- and S-type, in this condition (Figure 4.11.).

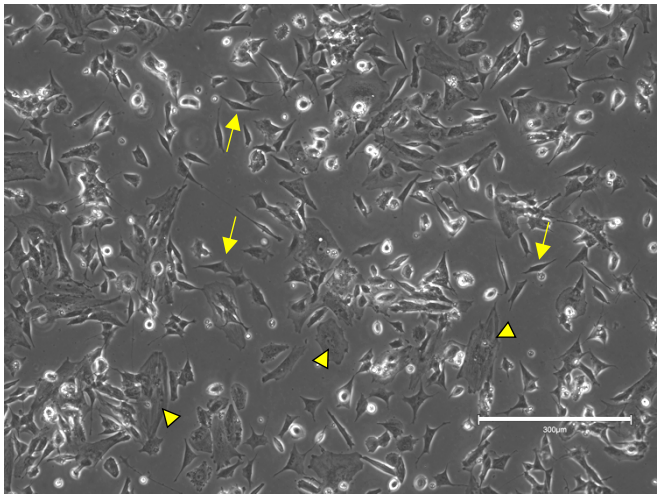


Figure 4.11. Phase contrast microphotograph of transduced SH-SY5Y cells, on the 6th day of selection with 5 µg/mL blasticidin. Both phenotypes of SH-SY5Y cells, N-type (more neuronal-like, arrows) and S-type (more epithelial-like, arrowheads), are visible. Bar = 300 µm.

After day 7 of selection, when only the transduced cells remained alive, these cells were plated to a 10 cm plate at a confluency of 50%, and we continued to maintain those cells under selective pressure. However, from this day on, the evolution of these cells started to get worse, and a greater rate of cell death could be perceived (Figure 4.12, day 17 of selection). On day 18 of selection, the conditioned culture medium was filtered to clean it from cellular debris, and 5 mL of these conditioned medium added to the cells. Several clusters of cells were visible in day 21, and cells were resuspended with trypsin and 2 mL of the filtered conditioned medium, and plated in a 12-well plate. In the following days and roughly every other day, half of the conditioned medium was replaced by fresh medium containing 5 µg/mL blasticidin, to get rid of some cellular debris found in the medium. On day 23 of selection, it was possible to observe some small clusters of adherent N-Type SH-SY5Y cells in the 12-well plate, which seemed to be slowly expanding in the following days (Figure 4.12.).

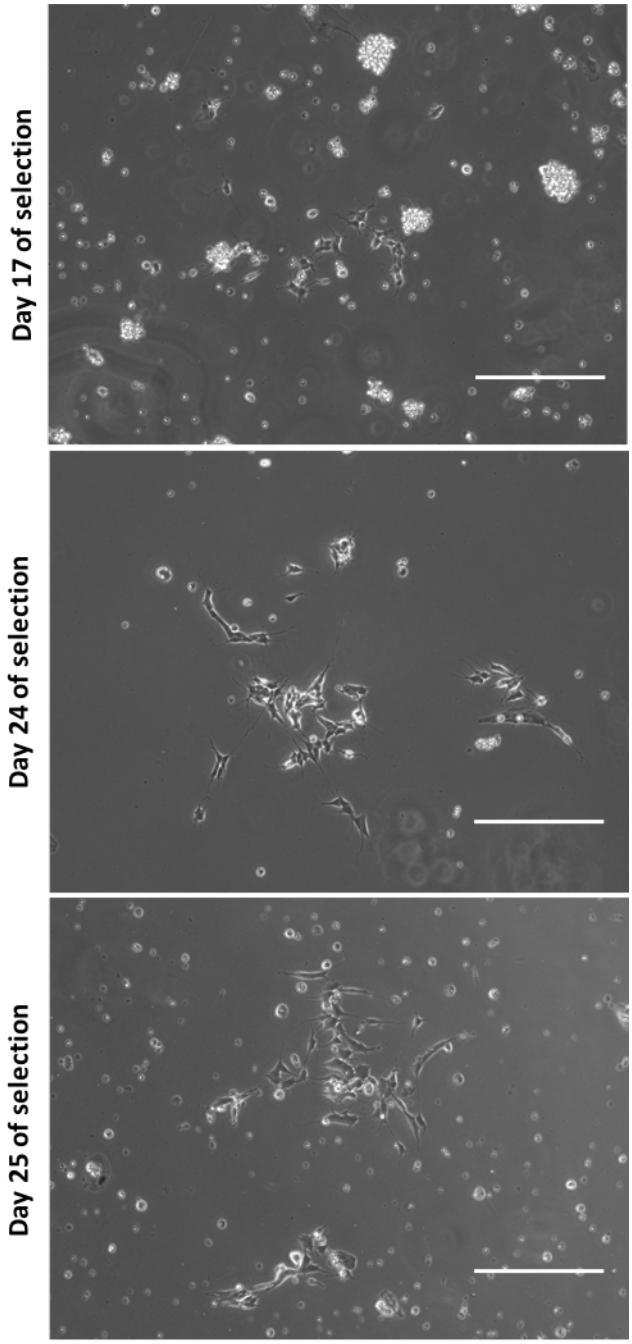


Figure 4.12. Long-term selection of SH-SY5Y-Cas9 cells at day 25. Phase contrast microphotographs show surviving cells with both phenotypes, potentially presenting the *SpCas9* gene integrated in their genome. Bar = 300 µm.

4.3. Bioinformatics analysis of the transcriptomes of partially regenerative SCI rodents

This section describes the results from a bioinformatics study comparing two transcriptomes of partially regenerative rodent SCI models. The initial objective was to perform an update on the initial bioinformatics study that led to the identification of *MPP3* and *RPRM*[37]. The aim beyond this was to find common robustly regulated genes, in an increased number of spinal cord partial regenerative paradigms. For that, a literature mining was first performed to search for recent transcriptomic studies reporting manipulations in rodent SCI models that improved spinal cord regeneration and functional recovery after SCI. As result, a fourth study was included to the initial set of three studies. However, this fourth study was only on a subacute time point post-injury (28 days post-injury, dpi), while all the first three included transcriptomics of acute time points (up to 1-week post-injury), when *MPP3* and *RPRM* were found deregulated. Further, the available transcriptomic data from the new study was already processed to differentially expressed genes (DEGs) obeying to previously set parameters. Given this, the fourth study was only compared to data of one of the original studies, of an approximate time point (20 dpi), processed in a similar way.

4.3.1. Selection of transcriptomic studies

The main aim of this analysis was to identify putative RAGs, genes whose expression is consistently regulated in various regenerative SCI paradigms. A search in Pubmed (https://pubmed.ncbi.nlm.nih.gov), using the keywords “spinal cord regeneration” OR “spinal cord repair” OR “transcriptome” AND “spinal cord injury”, with filter for *rattus* and *mus musculus* species, together with an additional search in GEO (https://www.ncbi.nlm.nih.gov/gds), Array Express (https://www.ebi.ac.uk/arrayexpress/) and SRA (https://www.ncbi.nlm.nih.gov/sra) databases, resulted in 12 studies (Supplementary table 1 in Annex). This set was further filtered. To fulfil the purposes of this bioinformatic work, the studies of interest were the ones where a potentially pro-regenerative treatment had been applied, with at least partial success, to the SCI animal model, and where a comparative transcriptomic study of the expression pattern at the lesion site had been performed. These selection criteria narrowed down the list to four studies (Table 4.1.), for further analysis and comparison. Three of these were the ones included in the original study[37]. The fourth study, PMC6854783 (Zeng, H. et al), was the one introduced in this update.

Table 4.1. Information and characteristics of the selected transcriptomic studies of rodent spinal cord injury models with regenerative approaches.

	E-GEOD-69334	**E-GEOD-34430**	**GSE62760**	**PMC6854783**
Our internal ref. and article title	Transcriptome analyses reveal molecular mechanisms underlying functional recovery after spinal cord injury	Differential gene expression in the EphA4 knockout spinal cord and analysis of the inflammatory response following spinal cord injury	T cell deficiency in spinal cord injury: altered locomotor recovery and whole-genome transcriptional analysis	Transcriptomic analysis of α-synuclein knockdown after T3 spinal cord injury in rats
Author(s)	Duan H, Ge W, Zhang A, Xi Y, Chen Z, Luo D, Cheng Y, Fan KS, Horvath S, Sofroniew MV, Cheng L, Yang Z, Sun YE, Li X.	Munro KM, Perreau VM, Turnley AM.	Satzer D, Miller C, Maxon J, Voth J, DiBartolomeo C, Mahoney R, Dutton JR, Low WC, Parr AM	Zeng H, Yu B, Yang Y, Xing H, Liu X, Zhou M
Year and DOI	2015 https://doi.org/10.1073/pnas.1510176112	2012 https://doi.org/10.1371/journal.pone.0037635	2015 10.1186/s12868-015-0212-0	2019 10.1186/s12864-019-6244-6
Database and dataset ID	Array Express, E-GEOD- 69334	Array Express, E-GEOD- 34430	GEO, GSE62760	DEGs taken from the article. RNA-Seq raw data was deposited in the NCBI Sequence Read Archive (SRA). Accession ID for Rattus norvegicus BioProject = PRJNA552942
Species	*Rattus norvegicus*	*Mus musculus*	*Rattus norvegicus*	*Rattus norvegicus*
Animals sex and age	Females 6- to 8-wk-old	Mix 10- to 14-wk-old	Females 16- to 18-wk-old	Male; Adults
Type and site of injury	Complete transection, T7-T8	Left hemisection whith complete transection, T13-L1	Contusion at force of kdyn, T8-T9	Contusion at force of 400 kdyn, T3
Tissue examined	Lesion site, Caudal, Rostral	Lesion site	Lesion site, Caudal, Rostral	Lesion site (5 mm above and below the T3 injury site)
Time points examined	1, 3, 10, 20, 30, 60, 90 dpi	4 dpi	1 wpi, 8 wpi	28 dpi
Treatment	NT3 treatment with chitosan tubular scaffold	EphA4 knockout mice	T-cell deficient athymic nude rats	Lentiviral vector-mediated α-Syn knockdown
Analysis	Affymetrix GeneChip (Rat Genome 230) 2.0	Affymetrix Mouse All-Exon ST Array GeneChipsTM	HiSeq 2000 (Illumina)	HiSeq xten (Illumina)
Statistics	Linear models and empirical Bayes methods with Limma, a R/Bioconductor software package, to assess differential expression and identify the top 10,000 transcripts, 7,500 of which were annotated genes.	Partek was used for gene expression analyses, with a p-value cut-off of 0.01 and a FC cut-off of 1.2	Transcriptional datasets for each time point were pooled using CuffMerge. Differences between strains at each time point were identified with CuffDiff. Gene expression differences with p-adjust < 0.05 were considered to be statistically significant	Comparisons between the two groups with Student's t-test or the Mann-Whitney test (as appropriate) Statistical significance of the DEGs: p-adjust< 0.05 &|log2FC| ≥ 1
Main conclusions	NT3-chitosan enhances new neurogenesis and angiogenesis, and reduces inflammatory responses	Subtle alterations in the neuroinflammatory response in injured EphA4 knockout spinal cords may contribute to the regeneration and recovery following SCI	AN rats have delayed tissue death and limit long-term recovery. T cell inhibition combined with other neuroprotective treatment may thus be a promising therapeutic avenue	Knockdown of α-Syn after SCI enhances motor function and promotes neurogenesis probably through enhancing cholinergic signaling pathways and neuroreceptor interactions

4.3.2. Data extraction

Files of the samples of interest were downloaded from the transcriptomic databases mentioned or from supplementary information of the scientific paper reporting the study:

For the transcriptomic study E-GEOD-69334, datasets of 3 samples per condition were extracted. These included data from the lesion site of injured NT3 treated animals, as well from lesioned animals that did not receive treatment, used as control samples, at the same time-point (20 dpi). An initial comparison analysis was performed with the 30 dpi dataset, but no common DEG was found.

Regarding PMC6854783, two Excel files were directly retrieved from the scientific paper. These corresponded to the lists of down-regulated and up-regulated DEGs, resulting from the comparison of the expression data from injured wild-type animals ('SCI group') and the data from injured α-Syn knockdown animals ('LV_SCI group'; LV from 'lentivirus'), upon defining a p-adjust<0.05 and a $|log2FC| \geq 1$. DEGs from this study could only be compared with E-GEOD-69334 data if the same filters used to retrieve these DEGs were applied to the E-GEOD-69334 dataset.

4.3.3. Data pre-processing and DEGs screening (E-GEOD-69334 20 dpi dataset)

Fold increases of the E-GEOD-69334 transcriptomic data of treated injured animals over the ones of injured non-treated ones, at 20 dpi, were calculated using Reactome (https://reactome.org/). The number of DEGs obtained in the E-GEOD-69334 study when p-adjust<0.05 and a $|log2FC| \geq 1$ were set as threshold values (as for the and in PMC6854783 study) are shown in Table 4.2.

Table 4.2. Number of identified up-regulated (UR) and down-regulated (DR) DEGs, with p-adjust<0.05 and a $|log2FC| \geq 1$, found in the E-GEOD-69334 and PMC6854783 studies.

Study	UR DEGs, p-value < 0.05	DR DEGs, p-value < 0.05
E-GEOD-69334	221	224
PMC6854783	209	127

4.3.4. Conversion to human orthologs

Both studies were performed in *R. norvegicus*, but we aimed to be able to compare findings to the ones from the original study, where DEGs were converted to human orthologs. The four DEGs' lists (Table 4.2) were uploaded to the BioMart online tool (https://www.ensembl.org/biomart/martview/), and human orthologs retrieved. Results for 'many2many' homology were excluded, as well as genes with human orthology confidence of 0.

For the E-GEOD-69334 study, 84.2% of the up-regulated genes and 74.6% of the down-regulated genes have a human ortholog. Regarding the PMC6854783 study, 88.5% and 67.7% human orthologs were retrieved for the up- and down-regulated genes, respectively. This information is resumed in Table 4.3.

Table 4.3. Number and percentage of UR and DR DEGs (resulting from the established parameters) for which human orthologs were found, per study.

Study	Human orthologs found for UR DEGs	Human orthologs found for DR DEGs
E-GEOD-69334	186 out of 221 – 84.2%	167 out of 224 – 74.6%
PMC6854783	185 out of 209 – 88.5%	86 out of 127 – 67.7%

4.3.5. Comparison of expression profiles

A Venn diagram was constructed to assess overlaps between deregulated DEGs, based on a p-adjust< 0.05 and a |log2FC| ≥ 1, in the two E-GEOD-69334 and PMC6854783 studies (Figure 4.13.).

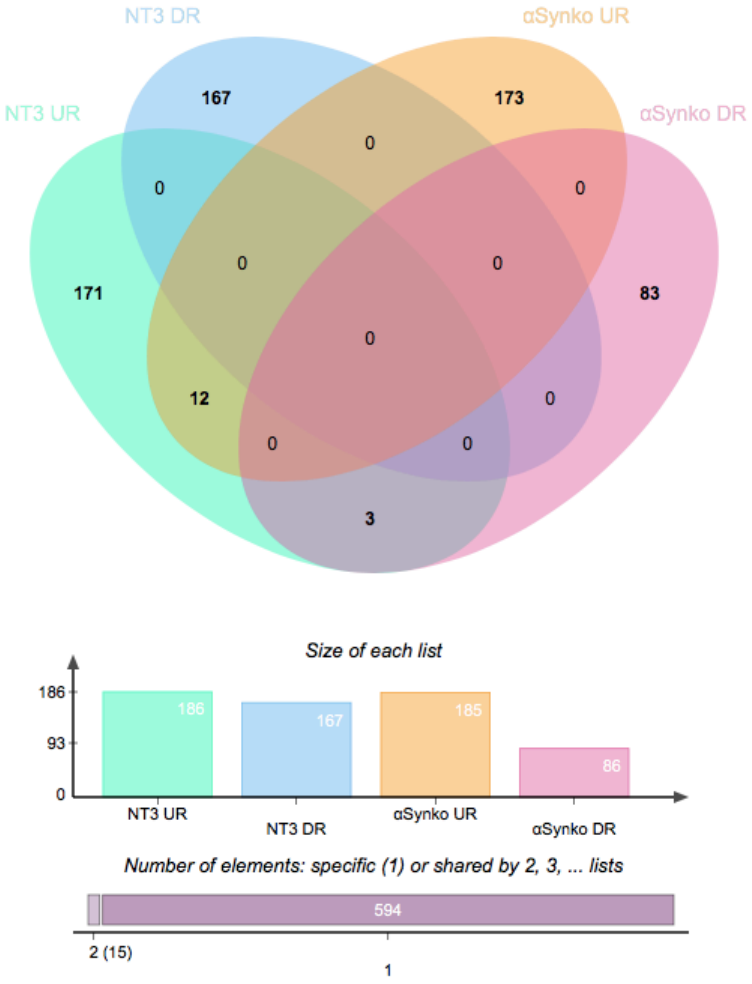


Figure 4.13. Comparison of expression profiles of the two analyzed studies. Venn diagram of the 4 analyzed datasets of up- and down-regulated DEGs. 12 differential expressed genes are commonly up-regulated in the two studies, and 3 of the up-regulated DEGs in the E-GEOD-69334 study were found to be down-regulated in the PMC6854783. The Venn diagram was created with the jVenn software.

12 up-regulated DEGs were found to be common to both studies and 3 other genes were also found to be common, however, up-regulated in the E-GEOD-69334 and down-regulated in the PMC6854783 study. These common genes are identified in Table 4.4., together with the respective log2FC values calculated for each study.

4.1.6. PPI network analysis

The online tool String was used to create networks of protein interactions and to find the enriched pathways and GO terms in it (Figure 4.14.). Within the 12 genes commonly up-regulated in the two studies, only 2 connections were identified with String, with the same protein – ATP2B. This can associate to other two DEGs: SLC24A2 (connection based on gene neighborhood, text-mining, and co-expression), and CALCA (connection based on text-mining and experimentally determined). String's neighborhood score results from systematic all-against-all genome comparisons, searching for preceding genome rearrangements or gene modifications such as gene gains, losses and fusions[68]; the text-mining score results from the statistical analysis of the co-occurrence of gene/protein names in abstracts, from a large number of scientific texts[68]; the co-expression score is based on similar pattern of mRNA expression measured by DNA arrays and similar technologies, across a large number of gene expression datasets; the experimental score is derived from experimental data, such as, affinity chromatography and immunoprecipitation.

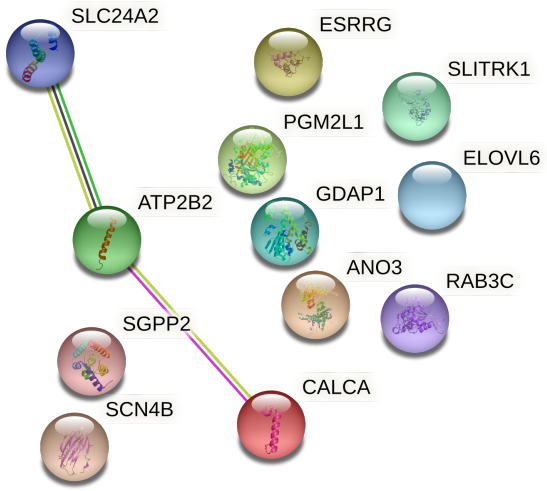


Figure 4.14. Protein interactions between up-regulated common genes. String's protein-protein interaction network based on Text-mining, Experiments, Databases, Co-expression, Neighborhood, Gene fusion and Co-occurrence data, with connection interaction scores of at least 0.4 and without expansion of the network. PPI created with the String software.

4.1.7. Gene Ontology and pathways enrichment analysis

To analyze the list of the overlapping up-regulated DEGs, UniProt and DAVID were first used to retrieve the pathways and GO terms associated to each gene. This is important because even if the genes do not interact directly with each other, they might contribute to some common pathways/functions, potentially involved in neuroregeneration. Results from this analysis are resumed in Table 4.4.

Overall, there are some GO terms and pathways that are associated to more than one gene from the list. Regarding GO terms, transmembrane transport (mainly of ions), transport and homeostasis of calcium, ATP binding, GTPase activity, activation of adenylate cyclase activity, G-protein signaling pathways, regulation of exocytosis, and inflammatory response. Additionally, more than one gene encodes for proteins that are integral components of the membrane or are located at synapses. Regarding to the KEGG and Reactome pathways, there is a predominance of the following signaling pathways: Calcium signaling, cAMP signaling and G-protein signaling.

Table 4.4. List of the common up-regulated genes between E-GEOD-69334 ('Study 1') and PMC6854783 ('Study 2'). Details about their physiological functions, Log2FC for each study, as well as the Gene Ontology (GO) terms and pathways associated to each gene are presented. Gene and Protein names, and protein code, were taken from Uniprot. In bold: terms and functions regarding the Nervous System. Some less relevant or repeated GO terms for the CALCA gene were here omitted due to space restrictions.

Gene & Protein names	Log2FC Study 1; Study 2	Known function	Gene ontology (biological process)	Gene ontology (molecular function)	Gene ontology (cellular component)	Pathways (Reactome and KEGGS)
ANO3 **ANO3, Anoctamin-3** **(Q9BYT9)**	4.023; 1.986	Appears to encode a Ca2+-gated chloride channel[69,70]. Related to **modulation of neuronal excitability**[69]. Mutations cause autosomal dominant **craniocervical dystonia**.[69,70]	Calcium activated galactosylceramide scrambling [GO:0061591]; calcium activated phosphatidylcholine scrambling [GO:0061590]; chloride transmembrane transport [GO:1902476]; detection of mechanical stimulus [GO:0050982]; detection of temperature stimulus [GO:0016048]; ion transmembrane transport [GO:0034220]; transmembrane transport [GO:0055085].	Chloride channel activity [GO:0005254]; phospholipid scramblase activity [GO:0017128]; protein dimerization activity [GO:0046983]	Integral component of membrane [GO:0016021]; plasma membrane [GO:0005886]	Reactome: Stimuli-sensing channels
ATP2B4 **PMCA4, Plasma membrane calcium-transporting ATPase 4** **(Q01814)**	1.416; 1.507	Encodes a Ca2+ pump, contributing for the intracellular calcium homeostasis[71]. **Modulates signaling pathways that regulate neuronal survival, differentiation, migration, and synaptogenesis**[72,73]. Is **associated with neural development**.	Calcium ion transport [GO:0006816]; cellular calcium ion homeostasis [GO:0006874]; ion transmembrane transport [GO:0034220]; **neuron differentiation** [GO:0030182]; regulation of cardiac conduction [GO:1903779]; regulation of cytosolic calcium ion concentration [GO:0051480]; sensory perception of sound [GO:0007605]	ATPase-coupled cation transmembrane transporter activity [GO:0019829]; ATP binding [GO:0005524]; calcium ion binding [GO:0005509]; calcium transmembrane transporter activity, phosphorylative mechanism [GO:0005388]; **calcium-transporting ATPase activity involved in regulation of presynaptic cytosolic calcium ion concentration** [GO:1905056]; calmodulin binding [GO:0005516]; metal ion binding [GO:0046872]; PDZ domain binding [GO:0030165]; protein C-terminus binding [GO:0008022].	Cytoplasm [GO:0005737]; extracellular exosome [GO:0070062]; **GABA-ergic synapse** [GO:0098982]; **glutamatergic synapse** [GO:0098978]; integral component of membrane [GO:0016021]; intracellular membrane-bounded organelle [GO:0043231]; plasma membrane [GO:0005886]; **postsynaptic density membrane** [GO:0098839]; **presynapse** [GO:0098793]	KEGG: Calcium signaling pathway; cGMP-PKG signaling pathway; cAMP signaling pathway; Adrenergic signaling in cardiomyocytes; Salivary secretion; Pancreatic secretion
CALCA **If CGRP-I, Calcitonin gene-related peptide 1** **(P06881)** **If Calcitonin** **(P01258)**	3.797; 2.560	CGRP-I Induces vasodilation. Its abundance in the CNS also points toward a **neurotransmitter or neuromodulator role**. **CGRP-I expression and release are increased by Nerve growth factor (NGF) in sensory neurons.**[74] **Its increased release from central terminals of sensory neurons augments nociception and contributes to the**	G protein-coupled receptor signaling pathway [GO:0007186]; cell-cell signaling [GO:0007267]; protein phosphorylation [GO:0006468]; receptor internalization [GO:0031623]; adenylate cyclase-activating G protein-coupled receptor signaling pathway [GO:0007189]; Activation of adenylate cyclase activity [GO:0007190]; activation of protein kinase activity [GO:0032147]; positive regulation of cytosolic calcium ion concentration involved in phospholipase C-activating G protein-coupled signaling pathway [GO:0051482]; **cellular response to nerve growth factor stimulus** [GO:1990090]; **neuropeptide signaling pathway** [GO:0007218]; cellular response to tumor necrosis factor	Calcitonin receptor binding [GO:0031716]; hormone activity [GO:0005179]; protein-containing complex binding [GO:0044877]; signaling receptor binding [GO:0005102]; identical protein binding [GO:0042802].	Cytoplasm [GO:0005737]; extracellular region [GO:0005576]; extracellular space [GO:0005615]; **neuronal cell body** [GO:0043025]; nucleus [GO:0005634]; **terminal bouton** [GO:0043195]	Reactome: G alpha (s) signalling events; Calcitonin-like ligand receptors; **ADORA2B mediated anti-inflammatory cytokines production**; **Amyloid fiber formation**

		development of central sensitization.[75] Calcitonin has a role in bone homeostasis, inhibiting osteoclastic activity (reduces calcium levels in the blood)[76].	[GO:0071356]; detection of temperature stimulus involved in **sensory perception of pain** [GO:0050965]; response to heat [GO:0009408]; response to pain [GO:0048265]; inflammatory response [GO:0006954]; positive regulation of: interleukin-1 alpha production [GO:0032730], interleukin-8 production [GO:0032757]; macrophage differentiation [GO:0045651]; leukocyte cell-cell adhesion [GO:0007159]; monocyte chemotaxis [GO:0002548]; innate immune response [GO:0045087]; antimicrobial humoral immune response mediated by antimicrobial peptide [GO:0061844]; **nervous system process involved in regulation of systemic arterial blood pressure** [GO:0001976]; vasculature development [GO:0001944]; endothelial cell migration [GO:0043542] and proliferation [GO:0001935]; regulation of blood pressure [GO:0008217]; negative regulation of: bone resorption [GO:0045779], ossification [GO:0030279], and osteoclast differentiation [GO:0045671].			
ELOVL6 **ELOVL6, Elongation of very long chain fatty acids protein 6** **(Q9H5J4)**	1.244; 1.566	Encodes a key enzyme in lipogenesis.[77] Also known as '**MASR, Myelin-associated SUR4 protein', present in myelinating Schwann cells, with a crucial role in normal myelin lipid synthesis**[78]	Fatty acid elongation, monounsaturated fatty acid [GO:0034625]; fatty acid elongation, polyunsaturated fatty acid [GO:0034626]; fatty acid elongation, saturated fatty acid [GO:0019367]; long-chain fatty acid biosynthetic process [GO:0042759]; long-chain fatty-acyl-CoA biosynthetic process [GO:0035338]; positive regulation of cold-induced thermogenesis [GO:0120162]; regulation of cholesterol biosynthetic process [GO:0045540]; sphingolipid biosynthetic process [GO:0030148]; unsaturated fatty acid biosynthetic process [GO:0006636]; very long-chain fatty acid biosynthetic process [GO:0042761].	3-oxo-arachidoyl-CoA synthase activity [GO:0102336]; 3-oxo-cerotoyl-CoA synthase activity [GO:0102337]; 3-oxo-lignoceronyl-CoA synthase activity [GO:0102338]; fatty acid elongase activity [GO:0009922]; very-long-chain 3-ketoacyl-CoA synthase activity [GO:0102756]	Endoplasmic reticulum [GO:0005783]; endoplasmic reticulum membrane [GO:0005789]; fatty acid elongase complex [GO:0009923]; integral component of endoplasmic reticulum membrane [GO:0030176]	KEGG: Fatty acid elongation, Biosynthesis of unsaturated fatty acids, Fatty acid metabolism. Reactome: Activation of gene expression by SREBF (SREBP); Synthesis of very long-chain fatty acyl-CoAs.
***ESRRG* (alias *ERR3*)** **ESRRG, Estrogen-related receptor gamma (alias ERR3)** **(P62508)**	1.593; 1.788	Encodes a receptor that binds specifically to an estrogen response element and activates reporter genes controlled by estrogen response elements. **Highly expressed in the nervous system, during embryogenesis and in adult brains**. **Present during neurogenesis**[79]**.** **Has a crucial role in neuronal differentiation;**	Positive regulation of cold-induced thermogenesis [GO:0120162]; positive regulation of transcription, DNA-templated [GO:0045893]; regulation of transcription, DNA-templated [GO:0006355]; regulation of transcription by RNA polymerase II [GO:0006357]; **retinoic acid receptor signaling pathway** [GO:0048384]; transcription initiation from RNA polymerase II promoter [GO:0006367]	AF-2 domain binding [GO:0050682]; DNA-binding transcription activator activity, RNA polymerase II-specific [GO:0001228]; DNA-binding transcription factor activity, RNA polymerase II-specific [GO:0000981]; identical protein binding [GO:0042802]; nuclear receptor activity [GO:0004879]; RNA polymerase II cis-regulatory region sequence-specific DNA binding [GO:0000978]; sequence-specific double-stranded DNA binding [GO:1990837]; steroid binding	Nuclear chromatin [GO:0000790]; nucleoplasm [GO:0005654]	Reactome: Nuclear Receptor transcription pathway

		its over-expression in SH-SY5Y cells leads to inhibition of cell proliferation and cell cycle arrest[80]**.**		[GO:0005496]; steroid hormone receptor activity [GO:0003707]; zinc ion binding [GO:0008270].		
GDAP1 **GDAP1, Ganglioside-induced differentiation-associated protein 1 (Q8TB36)**	3.138; 1.089	Involved in the regulation of mitochondrial dynamics and calcium homeostasis[81]. **Highly expressed in peripheral neurons**[82] Mutations may alter mitochondrial morphology and result in **peripheral neuropathy Charcot-Marie-Tooth disease**.[82]	Cellular response to vitamin D [GO:0071305]; mitochondrial fission [GO:0000266]; mitochondrial fusion [GO:0008053]; mitochondrion organization [GO:0007005]; protein targeting to mitochondrion [GO:0006626]; **response to retinoic acid** [GO:0032526]		Cytosol [GO:0005829]; integral component of mitochondrial outer membrane [GO:0031307]; membrane [GO:0016020]; mitochondrion [GO:0005739]; nucleus [GO:0005634]	Reactome: Class I peroxisomal membrane protein import
PGM2L1 **PGM2L, Gucose 1,6-bisphosphate synthase (Q6PCE3)**	1.188; 1.407	Synthesizes a cofactor for phosphomutases and putative regulator of glycolysis. **Expressed in brain where glucose-1,6-bisphosphate synthase activity is particularly high**.[83]	Canonical glycolysis [GO:0061621]; galactose catabolic process [GO:0019388]; glycogen biosynthetic process [GO:0005978]; glycogen catabolic process [GO:0005980]	Glucose-1,6-bisphosphate synthase activity [GO:0047933]; intramolecular transferase activity, phosphotransferases [GO:0016868]; metal ion binding [GO:0046872]	Cytosol [GO:0005829]	KEGG: Starch and sucrose metabolism. Reactome: Glycogen synthesis; Glycolysis; Glycogen breakdown (glycogenolysis); Galactose catabolism.
RAB3C **RAB3C, Ras-related protein Rab-3C (Q96E17)**	1.582; 1.423	Encodes a GTP-binding protein of the Ras superfamily **Modulation of secretory vesicle exocytosis**.[84,85]	Antigen processing and presentation [GO:0019882]; protein localization to plasma membrane [GO:0072659]; protein secretion [GO:0009306]; regulation of exocytosis [GO:0017157]; **vesicle docking involved in exocytosis** [GO:0006904]	GTPase activity [GO:0003924]; GTP binding [GO:0005525]; GTP-dependent protein binding [GO:0030742]; myosin V binding [GO:0031489]	Anchored component of synaptic vesicle membrane [GO:0098993]; cytoplasmic vesicle [GO:0031410]; cytosol [GO:0005829]; endosome [GO:0005768]; perinuclear region of cytoplasm [GO:0048471]; plasma membrane [GO:0005886]; synaptic vesicle [GO:0008021]; vesicle [GO:0031982].	Reactome: RAB geranylgeranylation
SCN4B **SCN4B, Sodium channel subunit beta-4 (Q8IWT1)**	2.071; 1.561	Encodes for the β4 protein, a subunit of voltage-gated sodium channels (NaV) in excitable tissues (generates action potentials); Has a role in regulating cellular functions such as migration, **differentiation** and **synaptic plasticity**.[86,87]	AV node cell action potential [GO:0086016]; cardiac muscle cell action potential involved in contraction [GO:0086002]; cardiac muscle contraction [GO:0060048]; membrane depolarization during cardiac muscle cell action potential [GO:0086012]; positive regulation of sodium ion transport [GO:0010765]; regulation of heart rate by cardiac conduction [GO:0086091]; regulation of sodium ion transmembrane transporter activity [GO:2000649]; regulation of ventricular cardiac muscle cell	Ion channel binding [GO:0044325]; sodium channel regulator activity [GO:0017080]; voltage-gated ion channel activity [GO:0005244]; voltage-gated sodium channel activity [GO:0005248]; voltage-gated sodium channel activity involved in cardiac muscle cell action potential [GO:0086006]	Intercalated disc [GO:0014704]; intrinsic component of plasma membrane [GO:0031226]; voltage-gated sodium channel complex [GO:0001518]	KEGG: Adrenergic signaling in cardiomyocytes. Reactome: Interaction between L1 and Ankyrins; Phase 0 - rapid depolarisation

			membrane repolarization [GO:0060307]; sodium ion transmembrane transport [GO:0035725]; sodium ion transport [GO:0006814].			
SGPP2 **SGPP2, Sphingosine-1-phosphate phosphatase 2 (Q8IWX5)**	3.014; 2.150	Encodes a protein that degrades the bioactive signaling molecule sphingosine 1-phosphate (S1P), which acts as an intracellular and extracellular signaling molecule, and regulates critical cell processes including apoptosis, cell growth, and immune function[88]; S1P signaling is associated with a variety of neuroinflammatory disorders, including multiple sclerosis and Alzheimer's disease.[89(p20)]	Phospholipid dephosphorylation [GO:0046839]; regulation of type B pancreatic cell proliferation [GO:0061469]; sphingolipid biosynthetic process [GO:0030148]; sphingosine metabolic process [GO:0006670]	Lipid phosphatase activity [GO:0042577]; sphingosine-1-phosphate phosphatase activity [GO:0042392]	Endoplasmic reticulum [GO:0005783]; endoplasmic reticulum membrane [GO:0005789]; integral component of membrane [GO:0016021]; membrane [GO:0016020]	KEGG: Sphingolipid metabolism, Sphingolipid signaling pathway. Reactome: Sphingolipid de novo biosynthesis
SLC24A2 **SLC24A2, Sodium/ potassium / calcium exchanger 2 (Q9UI40)**	1.876 1.682	Encodes an important K+-dependent Na+-Ca2+ exchanger that mediates **Ca2+ extrusion**; Overexpression of SLC24A2 reduced the expression of inflammatory cytokines in rodent models of sciatic nerve chronic constriction injury[35]	Calcium ion import across plasma membrane [GO:0098703]; calcium ion transmembrane transport [GO:0070588]; **cellular calcium ion homeostasis** [GO:0006874]; ion transmembrane transport [GO:0034220]; ion transport [GO:0006811]; learning [GO:0007612]; long-term synaptic depression [GO:0060292]; long-term synaptic potentiation [GO:0060291]; memory [GO:0007613]; potassium ion transmembrane transport [GO:0071805]; protein-containing complex assembly [GO:0065003]; response to stimulus [GO:0050896]; sodium ion transmembrane transport [GO:0035725]; visual perception [GO:0007601].	Calcium, potassium:sodium antiporter activity [GO:0008273]; calcium channel activity [GO:0005262]; symporter activity [GO:0015293]	Integral component of plasma membrane [GO:0005887]; photoreceptor inner segment [GO:0001917]; plasma membrane [GO:0005886]	Reactome: Sodium/Calcium exchangers
SLITRK1 **SLITRK1, SLIT and NTRK-like protein 1 (Q96PX8)**	1.841; 1.492	Regulates **synaptogenesis** between hippocampal neurons; Enhances **neuronal dendrite outgrowth**; It is localized in excitatory synapses and promotes their **differentiation**; Its overexpression in hippocampal neurons increased the number of synaptic contacts on these neurons.[90]	Adult behavior [GO:0030534]; **axonogenesis** [GO:0007409]; homeostatic process [GO:0042592]; multicellular organism growth [GO:0035264]; positive regulation of axonogenesis [GO:0050772]; positive regulation of **synapse assembly** [GO:0051965]; regulation of presynapse assembly [GO:1905606]; synapse assembly [GO:0007416]; synaptic membrane adhesion [GO:0099560].		Extracellular region [GO:0005576]; integral component of membrane [GO:0016021]; plasma membrane [GO:0005886]; synapse [GO:0045202]	Reactome: Receptor-type tyrosine-protein phosphatases

5. Discussion

In the present work, a review on the current gene and molecular therapies for SCI was made.

Molecular and gene targets for SCI therapy are focusing on key molecules known to be involved in neuroregeneration in other tissues or animals, such as Ephr4[39], to manipulate them in the CNS neurons and improve repair. Other approaches work on overcoming the inhibitory environment in the injury site, reducing inhibitory factors such as AGIs, or in promoting a pro-regenerative one, by supplementing neurotrophins to the injury site, to promote neuroregeneration.

Regarding gene therapies, besides the modulation of gene expression, other factors such as non-coding RNAs, may play a role in the development of the lesion and therefore, their potential for futures investigation should not be excluded. Gene networks involved in the axon regeneration process can be regulated not only by single or multiple differentially-expressed transcription factors, but also by a variety of regulatory RNAs, including microRNAs (miRNAs)[18]. Several miRNAs, such as miR-9, miR-219, and miR-384-5p, have been recognized to be differentially expressed after SCI and, thereby, may represent novel biomarkers for the diagnosis, treatment, and prognosis of such injuries[3].

Long non-coding RNAs (lncRNAs), which constitute a large type of noncoding RNAs, may also be of importance in SCI. LncRNAs have been emerging as important regulators of multiple disease processes, since there are increasing evidences that show they have multiple domains that can bind to DNA, RNA, and proteins. Through these, lncRNAs can modify the chromatin status, regulate transcription and translation, and ultimately affect gene expression. Its functional significance in SCI and the mechanisms underlying their involvement have not yet been completely decoded, although some studies in rodents after traumatic SCI have shown potential roles in cell proliferation, survival, migration, and genomic stability maintenance[3].

Circular RNAs (circRNAs) are another subtype of noncoding RNAs that can play a role in SCI. Emerging as novel RNA molecules, single-stranded with a covalently closed loop structure without 5' and 3' ends, circRNAs have potentially interesting biological functions, namely as genome regulators[5]. With the development of gene sequencing technology, namely microarray and subtractive hybridization analysis, circRNAs have been revealed to be implicated in various cellular and molecular mechanisms after CNS injuries, by interacting with DNA, messenger RNA (mRNA), or proteins. In the CNS, where they are particularly abundant, they seem to be associated with neuronal development and synaptic plasticity[5].

Nevertheless, despite the several efforts in diverse areas to develop therapies for SCI, molecular therapies are the main focus of research, and are still growing. However, even for these there is still much to be investigated to understand their efficacy in the treatment of this pathology, both *per se* and in combined therapy with other approaches, such as cellular and gene therapy.

To design a gene therapy for SCI, based on the modulation of gene expression, regeneration-associated genes (RAGs) must be identified. The number of RAGs known so far is high, and it would be difficult to modulate them all in one single therapy. It is thus necessary to identify a small set of 'hierarchically superior RAGs' that are and/or common genes consistently expressed during spinal cord regeneration. For this last, studies on spinal cord naturally regenerative animals (like zebrafish or tadpoles) can be

used[45], as well as studies of usually non-regenerative animals, like rodents, which were intervened genetically or through the implantation of a scaffold or the administration of a drug molecular therapy. From a previous analysis of three transcriptomic studies of SCI rodent models with partial regenerating spinal cords, in which such a treatment was tested, 68 differentially expressed genes (DEGs) were found in common. Two of these were here studied – *MPP3* and *RPRM*. The main goal of the work of this thesis was to develop molecular tools to study these genes in regenerative-related cellular processes, in neuroblastoma SH-SY5Y cells.

MPP3 belongs to the family of MAGUK proteins, that have important structural and signalling roles. At synaptic junctions, these proteins regulate synaptic development and the release of neurotransmitters from synaptic vesicles[16,41]. In the context of SCI, *MPP3* overexpression might contribute for the development and plasticity of synapses, conserving and improving neuronal circuits.

The *MPP3* mRNA was found up-regulated in the three pro-regenerative rodent studies, particularly in the first week post-injury[37]. So, gain-in-function studies were planned, where this gene would be overexpressed in neuroblastoma cells, under basal and differentiating conditions.

Unfortunately, we had no access to a full-length sequence coding for *MPP3* in a plasmid of eukaryotic expression. Nevertheless, we were given two plasmids that could be used to obtain a full length *MPP3* sequence in a eukaryotic expression vector. Part of the *MPP3* cDNA was subcloned from a prokaryotic to an eukaryotic expression vector, and the full *MPP3* CDS reconstructed. Positive colonies, that survived ampicillin selection, were monitored by 'colony-PCR', and one of the colonies transformed with recombinant pBABE fl-MPP3 expanded and used to amplify the recombinant plasmid. The sequence of the recombinant fl-MPP3 was confirmed by DNA sequencing. With the two primers used, we were not able to sequence the full *MPP3* CDS. Results so far indicate that the truncated region was reconstructed, as expected, and no mutation could be seen in the *MPP3* sequence. Nevertheless, a new 5' primer will be designed, and the remaining yet non-sequenced region will be sequenced. Further, an insertion was apparent in the sequence of the *MPP3* CMV promotor, but that part of the electrophoretogram was not very conclusive (two peaks very nearby, with some overlap).

To try to understand if the recombinant plasmid could express *MPP3* in the neuroblastoma cells, and if any effect was visible in the cells' phenotype, pBABE fl-MPP3 was transfected into SH-SY5Y cells, and overexpression analysed by Immunofluorescence techniques. Few transfected cells were observed, dead or apparently compromised. Since they were in a small number, it is possible that other successfully transfected cells had already died and were washed away before or during the proceedings of the Immunocytochemistry protocol. This means *MPP3* overexpression was most probably cytotoxic to the cell, with extra MPP3 molecules dysregulating its physiological pathways and consequently dysregulating and compromising other cellular functions.

Beta tubulin III was used to double label the cells in this transfection experiment, since it is a neuronal tubulin isoform, specifically localized to neurons, and its expression correlates with neuronal differentiation. Therefore, if MPP3 was implicated in some neuroregenerative processes such as differentiation and neuritogenesis, it should be possible to observe some correlation between the staining intensities of these two proteins, which was not observed, neither in the transfected nor in non-

transfected cells. In the future, we will repeat this experiment following confirmation of the recombinant fl-*MPP3* cDNA. Also, more physiological 'loss-of-function' MPP3 experiments will be designed, for example using CRISPR/Cas9 techniques.

RPRM, a transcriptional target for the tumour suppressor gene p53 that is mainly expressed in the CNS, inhibits the activation of the Cdc2/Cyclin B1 complex, essential for the cell to enter mitosis. Data from tumour studies suggest that when RPRM is transiently up-regulated not only it inhibits cell proliferation, but also impairs cell survival[91]. Regarding its potential role in neuroregeneration, its down-regulation in mitotic cells such as stem cells, is expected to allow cell proliferation and increase cell survival; in neurons, its down-regulation may contribute to improve neuronal survival of neurons affected or near the injured zone.

Since *RPRM* was found down-regulated in the previous bioinformatics analysis[37], as well as in another more recent work by the group that aims to unveil the early transcriptomic response of regenerative neurons, a knockout protocol was designed. The CRISPR/Cas9 gene editing tool and lentiviral particles (LVLPs) to assemble the system, were chosen as tools for this loss-of-function' experiment, due to their advantages described in the Introduction section.

As for the cells chosen to knockout *RPRM*, they were selected as they reflect the system under study, to ensure the results will have biological relevance. The goal is to search for potential neuroregenerative properties of the gene, including its effects on proliferation, cell survival and neuronal differentiation. It is important to keep in mind that every cellular model has advantages and limitations and therefore, a range of factors must be considered before choosing one, namely the *in vivo* relevance and human representation, the genome, the ease of manipulation, among others. Since *RPRM* was found downregulated in non-mitotic cells (neurons) but potentially also in mitotic cells at the lesioned tissue, including stem cells, we decided it would be interesting and relevant to study its effects in different cellular models: in a human neuroblastoma cell line (non-differentiated and under differentiation); in neural stem cells (non-differentiated and under differentiation); and in differentiating rat primary neurons.

In this study we have only used the human SH-SY5Y neuroblastoma cell line. These cells are a known model of neuronal differentiation, commonly used in our group[92]. Hence, it was decided to firstly create a stable SH-SY5Y cell line expressing *SpCas9*, so that in the future this system is ready to be used in these cells. To do so, it was decided to use LVs, since infection with these viruses leads to the integration of DNA into the genome of the host cell, which means it will be stably expressed together with the original cellular genome (if no other factors lead to its exclusion from the genome). A stable expression, contrary to a transient one, will allow the future use of these cells expressing the endonuclease to edit other genes potentially associated to neuroregeneration.

Firstly, it was necessary to produce LVLPs containing the DNA of interest, in this case, the *SpCas9* gene. The most used cell lines for LVLPs production are HEK293 and HEK293T cells. HEK293T might be preferred due to increased cell growth and higher transfection efficiency. Additionally, they contain the SV40 T-antigen, which renders them more efficient for vector production[56]. The protocol here used

to produce lentiviral particles encoding SpCas9, was successful: only the SH-SY5Y cells transduced with the filtered HEK293T conditioned media resisted to the selection media containing blasticidin. This confirmed the delivery of LVLPs containing the pLentiCas9-BLAST to the target neuroblastoma cells.

However, part of this protocol – the selection of stable SH-SY5Y expressing *SpCas9* – may need to be optimized. The selection of these cells with 5 µg/mL of blasticidin started two days after their transduction and, by monitoring the cells over a period of one week we observed what we expected. The negative non-transduced control cells that, consequently, did not acquire resistance genes to the blasticidin antibiotic, progressively started to die. In the well in which the LVLPs were added, despite some initial cell death (also expected, since the transduction efficiency is not 100%), many cells remained adherent. Unexpectedly, after 7 days of selection, the cells of the transduction condition, that until then had survived and were well adhered to the plate, also started to die or change morphologically. Several hypotheses that emerged for this event were: 1) In those first days of selection, SH-SY5Y cells expressing the transduced vector as a episome (not integrated into the genome) were also being selected; however, after losing the episomal lentiviral transgenes, they also lost the resistance to blasticidin; 2) Cells transduced and with the *SpCas9* and the blasticidin resistance gene integrated in the genome were destroying the antibiotic, and neighbour non-transduced cells could survive for some days; 3) The *SpCas9* and gene resistance construct was indeed inserted in the host genome, but interfered with the expression of essential proteins, compromising the survival of the cells; 4) Silencing of the internal promoter of the construct, used to express the transgenes, occurred by methylation, leading to loss of expression.

Currently, SH-SY5Y-Cas9 are still being selected. One approach that in the future can be tested to avoid the selection of cells with episomal expression of the transgene or hypothesis 2 (on the surviving non-transduced neighbour cells), consists in starting the selection with a higher concentration of the antibiotic (increasing the selective pressure of the medium) and later reduce it.

Since the main goal is to obtain a stable cell line of SH-SY5Y expressing *SpCas9* for future experiments, it is essential to work with a clonal population rather than a polyclonal one, in which cells have integrated the transgene in distinct regions of the genome and express it in different amounts. Thereby, after obtaining a greater confluence of our polyclonal transduced cells under selection, it will follow a protocol for single-cell cloning, to find single cell clones and produce clonal populations from them, where the cells will be genetically identical. Upon characterization using qRT-PCR, a clone that shows good *SpCas9* expression will be selected, and used to knockout RPRM by CRISPR/Cas9 means. The RPRM knockout cells will be used in proliferation and neuronal differentiation assays.

The last specific objective of this thesis was to extend the original **bioinformatics analysis of potential RAGs**, which can be later modulated by CRISPR/Cas9 techniques in our SH-SY5Y-Cas9 cells. The main goal here was to identify common DEGs between two studies in rat models in which distinct potential therapeutic approaches were tested to promote neuroregeneration. One of the original studies was compared to a more recent study, that used lentivirus to deliver Snca-shRNA into the lesioned spinal cords of rats, to down-regulate α-synuclein, well know due to its central role in Parkinson's disease.[93] The authors defended that the improved recovery was due to improved motor function and

promotion of neurogenesis via an enhancement of cholinergic signalling pathways and neuroreceptor interactions. This study had transcriptomic data from a later post-injury time (28 dpi) and was thus compared to one of the original studies that presented transcriptomic data of the lesion site, at similar time points post-injury.

Bioinformatic analysis can be an important tool to unveil the more relevant biological pathways related to the neuroregenerative approaches used in these studies. This type of analysis, often used as a starting point for functional *in vitro* studies, needs to be carefully planned, so that their conclusions are valid to allow that translation.

Firstly, the identification of DEGs was based on quite restrictive parameters (compared to most similar studies that performed the same type of analysis): p-adjust <0.05 and | log2FC | ≥ 1. These were the selection parameters set by authors of the PMC6854783 study, from which we had only available the lists of up- and down- regulated DEGs. These parameters naturally reduce the list of DEGs found in each study and, as expected, the DEGs common to both studies. However, on the other hand, this allowed us to identify with a high degree of confidence genes that present high differences in expression between control and treatment conditions. Consequently, this offers more confidence that the DEGs common to the two different therapeutic approaches may be essential for regeneration of the spinal cord. Twelve up-regulated genes were identified as commonly up-regulated in both studies (no common down-regulated gene was observed). Possible interactions between proteins encoded by these genes were sought using String and were based on Text-mining, Experiments, Databases, Co-expression, Neighbourhood, Gene fusion and Co-occurrence data, with connection interaction scores of at least 0.4 and without expansion of the network. As the short list of common DEGs was already, in itself, restrictive, it was decided not to further restrict possible interactions among the proteins encoded by those DEGs. By selecting an interaction score of medium confidence (0.400), as well as all the available options of sources of interaction, some interesting or unexpected interactions could arise, some that have not been proven experimentally or associated with spinal cord injury, for example.

Only 1 protein, **ATP2B**, was retrieved as potentially binding to two other proteins, SLC24A2 and CALCA, and these two do not interact with each other. *ATP2B* encodes a Ca^{2+} pump that contributes for the intracellular calcium homeostasis[71] and was shown to modulate signalling pathways involved in neuronal survival, differentiation, migration, and synaptogenesis[72,73] . These pathways clearly contribute for neuroregeneration following a spinal cord injury.

SLC24A2 encodes a K^+-dependent Na^+-Ca^{2+} exchanger that mediates Ca^{2+} extrusion, thereby contributing for the calcium homeostasis, such as ATP2B does. Additionally, an experimental study in rodent models has demonstrated an anti-inflammatory role for SLC24A2, by reducing the expression of inflammatory cytokines[35]. Likewise, the **SLC24A2** protein has an important role in the regulation of Ca^{2+} extrusion in cellular environments with dynamic Ca^{2+} fluxes, like it is the case of neuronal synapses.[94] Its overexpression, as for the other above indicated genes, might help decreasing the excessive intracellular Ca^{2+} levels, involved in neuronal apoptosis.[95]

CALCA encodes for the Calcitonin protein, involved in calcium homeostasis[76] or, by alternative splicing, the Calcitonin gene-related peptide 1 (CGRP). This later is highly abundant in the CNS, where it appears

to have a neurotransmitter or neuromodulator role. In fact, it is the main neurotransmitter of the nociceptive sensory c fibers.[96] It is thought that CGRP operates through the intracellular molecules Protein kinase C (PKC), PKA, and MAP kinase (MAPK), involved in signalling pathways that mediate neuritogenesis.[97] It has also been related to cytokine production, but when observing the context of its up-regulation together with *ATP2B* and *SLC24A2*, we believe that this would not be the case here.

All these three genes have roles in the calcium homeostasis maintenance, with calcium being important at physiological levels in many signalling events, but with excessive calcium levels in some neurological conditions being cytotoxic and usually a signal of the imminent death of neurons and axons. In the pathophysiology of the spinal cord injury, excessive calcium can also have a detrimental role as it promotes pro-reactive transcriptional changes and consequent activation of resting astrocytes.[98] The up-regulation of these three genes together, suggests that they may function synergistically to maintain intracellular calcium homeostasis and promote signalling pathways involved in neuronal survival, differentiation, and synaptogenesis, as observed for the protein coded by *ATP2B.*[72,73]

Further, another gene coding for protein involved in calcium homeostasis, was also commonly up-regulated. ***GDAP1*** is highly expressed in peripheral neurons, where the GDAP1 protein is involved in the regulation of mitochondrial dynamics and calcium homeostasis[81], both extremely affected by the SCI. Overexpression of this gene may help to restore the ability of mitochondria to maintain homeostasis, and consequently to reduce oxidative stress and excessive detrimental intracellular calcium levels.[99]

Other genes observed to be up-regulated in both studies, have also interesting roles or were already associated to functions potentially relevant for neural tissue repair.

The protein coded by ***PGM2L1*** is a glucose 1,6-bisphosphate synthase whose activity is particularly high in the brain. Involved in Glycogen synthesis and Glycolysis pathways, it is important in a metabolic active regenerating tissue.

SGPP2 encodes a protein that degrades the signaling molecule sphingosine 1-phosphate (S1P), which regulates cellular processes like apoptosis, cell growth, and immune function[88], all important in repair following SCI. Further, the regulation of the S1P signaling is highly important in a neuropathology associated to inflammation, such as SCI, since this signaling has been associated to neuroinflammatory disorders, including multiple sclerosis and Alzheimer's disease.[89(p20)]

ESRRG is associated with the retinoic acid receptor signalling pathway, a morphogen very important during the development of the nervous system, and used in SH-SY5Y neuronal differentiation.[92] Its overexpression may induce cell cycle arrest and inhibit cell proliferation[80], important for the onset of neuronal differentiation.

SLITRK1 encodes a transmembrane protein, that enhances dendrite outgrowth and localizes in excitatory synapses, promoting their development. Its overexpression in hippocampal neurons has shown to increase the number of synaptic contacts on these neurons.[90] Therefore, overexpression of this gene is important for dendritogenesis and for synaptic maturation, key in neuroregeneration, and

may help to improve neuronal synaptic transmission and plasticity, essential to support the neuronal function.

SCN4B encodes for the β4 protein, a subunit of voltage-gated sodium channels (NaV) in excitable tissues. It is involved in neuronal differentiation, synaptic plasticity and in the generation of action potentials.[86,87]

ELOVL6 is an enzyme involved in the elongation of very long fatty acids and has a crucial role in the myelin lipid synthesis in Schwann cells (from the PNS).[78] It is thus very likely that it might have a similar role in the oligodendrocytes of the CNS, necessary to myelinate the axonal fibres of the spinal cord circuits. If this is true, its overexpression would improve axonal myelination in SCI and, consequently, the conductive capacity of surviving axons and the 're-wiring' of the spinal cord.

RAB3C encodes a GTP-binding protein of the Ras superfamily, involved in the modulation of secretory vesicle exocytosis. In synaptic vesicles, it has a role in neurotransmitter release[84], and is thus important for the function of newly regenerated neurons.

Mutations in the ***ANO3*** gene cause autosomal dominant craniocervical dystonia.[69,70] Dystonia is a neurological movement disorder syndrome believed to be caused by a pathology of the CNS related to motor function control. The coded Anoctamin-3 protein appears to be a Ca^{2+}-gated chloride channel[69,70] that modulates neuronal excitability[69]. Hence, correct Ca^{2+} concentrations are important for this protein channel to modulate the correct neuronal response to electric impulses.

In **summary**, and as above indicated, further studies are needed to enlighten the putative role of *MPP3* and *RPRM* on cell proliferation, cell survival and neuronal differentiation. For this purpose, following work will include modulation of *MPP3* and RPRIMO genes for 'loss-of-function' assays related to neuronal proliferation, survival and neuronal differentiation, potentially relevant for future gene therapy in SCI patients. These assays should be carried not only on non-differentiated SH-SY5Y cells, but also, in the case of differentiation and survival, on already differentiated SH-SY5Y cells (phenotype closer to neuronal cells) and, further ahead, on neural stem cells and rat primary neurons. Additionally, the bioinformatics comparative study revealed the high importance of regulating calcium homeostasis at a subacute time point of recovery following a traumatic injury to the spinal cord. This is worth pursuing in following functional *in vitro* assays and further on in *in vivo* assays, by testing a multiple gene or molecular SCI therapy in rodent SCI models.

6. References

1. Hutson TH, Di Giovanni S. The translational landscape in spinal cord injury: focus on neuroplasticity and regeneration. *Nat Rev Neurol*. 2019;15(12):732-745. doi:10.1038/s41582-019-0280-3

2. Hachem LD, Ahuja CS, Fehlings MG. Assessment and management of acute spinal cord injury: From point of injury to rehabilitation. *J Spinal Cord Med*. 2017;40(6):665-675. doi:10.1080/10790268.2017.1329076

3. Wang F, Liu J, Wang X, et al. The Emerging Role of lncRNAs in Spinal Cord Injury. *BioMed Res Int*. 2019;2019:1-9. doi:10.1155/2019/3467121

4. Luo S, Xu H, Zuo Y, Liu X, All AH. A Review of Functional Electrical Stimulation Treatment in Spinal Cord Injury. *NeuroMolecular Med*. 2020;22(4):447-463. doi:10.1007/s12017-019-08589-9

5. Qu X, Li Z, Chen J, Hou L. The emerging roles of circular RNAs in CNS injuries. *J Neurosci Res*. 2020;98(7):1485-1497. doi:10.1002/jnr.24591

6. Jin MC, Medress ZA, Azad TD, Doulames VM, Veeravagu A. Stem cell therapies for acute spinal cord injury in humans: a review. *Neurosurg Focus*. 2019;46(3):E10. doi:10.3171/2018.12.FOCUS18602

7. van den Berg MEL, Castellote JM, Mahillo-Fernandez I, de Pedro-Cuesta J. Incidence of Spinal Cord Injury Worldwide: A Systematic Review. *Neuroepidemiology*. 2010;34(3):184-192. doi:10.1159/000279335

8. Bloom O, Herman PE, Spungen AM. Systemic inflammation in traumatic spinal cord injury. *Exp Neurol*. 2020;325:113143. doi:10.1016/j.expneurol.2019.113143

9. Tang R, Botchway BOA, Meng Y, et al. The Inhibition of Inflammatory Signaling Pathway by Secretory Leukocyte Protease Inhibitor can Improve Spinal Cord Injury. *Cell Mol Neurobiol*. 2020;40(7):1067-1073. doi:10.1007/s10571-020-00799-1

10. Musselman KE, Shah M, Zariffa J. Rehabilitation technologies and interventions for individuals with spinal cord injury: translational potential of current trends. *J NeuroEngineering Rehabil*. 2018;15(1):40. doi:10.1186/s12984-018-0386-7

11. Badhiwala JH, Ahuja CS, Fehlings MG. Time is spine: a review of translational advances in spinal cord injury. *J Neurosurg Spine*. 2019;30(1):1-18. doi:10.3171/2018.9.SPINE18682

12. Caruso MC, Daugherty MC, Moody SM, Falcone RA, Bierbrauer KS, Geis GL. Lessons learned from administration of high-dose methylprednisolone sodium succinate for acute pediatric spinal cord injuries. *J Neurosurg Pediatr*. 2017;20(6):567-574. doi:10.3171/2017.7.PEDS1756

13. Bracken MB, Holford TR. Effects of timing of methylprednisolone or naloxone administration on recovery of segmental and long-tract neurological function in NASCIS 2. *J Neurosurg*. 1993;79(4):500-507. doi:10.3171/jns.1993.79.4.0500

14. Bledsoe BE, Wesley AK, Salomone JP. High-dose steroids for acute spinal cord injury in emergency medical services. *Prehosp Emerg Care*. 2004;8(3):313-316. doi:10.1080/312704000322

15. Hirokawa T, Zou Y, Kurihara Y, et al. Regulation of axonal regeneration by the level of function of the endogenous Nogo receptor antagonist LOTUS. *Sci Rep*. 2017;7(1):12119. doi:10.1038/s41598-017-12449-6

16. Liu Y, Wang X, Li W, et al. A sensitized IGF1 treatment restores corticospinal axon-dependent functions. Published online 2018:34.

17. Griffin JM, Fackelmeier B, Clemett CA, et al. Astrocyte-selective AAV-ADAMTS4 gene therapy combined with hindlimb rehabilitation promotes functional recovery after spinal cord injury. *Exp Neurol*. 2020;327:113232. doi:10.1016/j.expneurol.2020.113232

18. Dulin JN, Antunes-Martins A, Chandran V, et al. Transcriptomic Approaches to Neural Repair. *J Neurosci*. 2015;35(41):13860-13867. doi:10.1523/JNEUROSCI.2599-15.2015

19. Schwab ME, Strittmatter SM. Nogo limits neural plasticity and recovery from injury. *Curr Opin Neurobiol*. 2014;27:53-60. doi:10.1016/j.conb.2014.02.011

20. Darian-Smith C. Synaptic Plasticity, Neurogenesis, and Functional Recovery after Spinal Cord Injury. *The Neuroscientist*. 2009;15(2):149-165. doi:10.1177/1073858408331372

21. National Academy of Medicine, National Academies of Sciences, Engineering, and Medicine, Committee on Human Gene Editing: Scientific, Medical, and Ethical Considerations. A, The Basic Science of Genome Editing. In: *Human Genome Editing: Science, Ethics, and Governance.* National Academies Press (US); 2017. https://www.ncbi.nlm.nih.gov/books/NBK447276/

22. Lino CA, Harper JC, Carney JP, Timlin JA. Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv*. 2018;25(1):1234-1257. doi:10.1080/10717544.2018.1474964

23. Merck. Merck Manual for Gene Editing 101; 2020. https://www.sigmaaldrich.com/life-science/functional-genomics-and-rnai/crispr-cas9-gene-editing/ebook.html

24. Xu CL, Ruan MZC, Mahajan VB, Tsang SH. Viral Delivery Systems for CRISPR. *Viruses*. 2019;11(1):28. doi:10.3390/v11010028

25. Dickinson DJ, Goldstein B. CRISPR-Based Methods for Caenorhabditis elegans Genome Engineering. *Genetics*. 2016;202(3):885-901. doi:10.1534/genetics.115.182162

26. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. 2012;337:816-821. doi:10.1126/science.1225829

27. Collias D, Leenay RT, Slotkowski RA, et al. A positive, growth-based PAM screen identifies noncanonical motifs recognized by the S. pyogenes Cas9. *Sci Adv*. 2020;6(29):eabb4054. doi:10.1126/sciadv.abb4054

28. Karvelis T, Gasiunas G, Young J, et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol*. 2015;16(1):253. doi:10.1186/s13059-015-0818-7

29. Lin W, Xu L, Lin S, et al. Characterisation of multipotent stem cells from human peripheral blood using an improved protocol. *J Orthop Transl*. 2019;19:18-28. doi:10.1016/j.jot.2019.02.003

30. Goswami R, Subramanian G, Silayeva L, et al. Gene Therapy Leaves a Vicious Cycle. *Front Oncol*. 2019;9:297. doi:10.3389/fonc.2019.00297

31. Paschon V, Correia FF, Morena BC, et al. CRISPR, Prime Editing, Optogenetics, and DREADDs: New Therapeutic Approaches Provided by Emerging Technologies in the Treatment of Spinal Cord Injury. *Mol Neurobiol*. 2020;57(4):2085-2100. doi:10.1007/s12035-019-01861-w

32. Song YH, Agrawal NK, Griffin JM, Schmidt CE. Recent advances in nanotherapeutic strategies for spinal cord injury repair. *Adv Drug Deliv Rev*. 2019;148:38-59. doi:10.1016/j.addr.2018.12.011

33. Zavvarian M-M, Toossi A, Khazaei M, Hong J, Fehlings M. Novel innovations in cell and gene therapies for spinal cord injury. *F1000Research*. 2020;9:279. doi:10.12688/f1000research.21989.1

34. Lou WP-K, Wang L, Long C, Liu L, Fei J-F. Direct Gene Knock-out of Axolotl Spinal Cord Neural Stem Cells via Electroporation of CAS9 Protein-gRNA Complexes. *J Vis Exp*. 2019;(149):59850. doi:10.3791/59850

35. Zhou X, He H, Wang P. A critical role for miR-135a-5p-mediated regulation of SLC24A2 in neuropathic pain. *Mol Med Rep*. 2020;22(3):2115-2122.

36. Hu Y, Liu S, Zhu B-M. CRISPR/Cas9-Induced Loss of Keap1 Enhances Anti-oxidation in Rat Adipose-Derived Mesenchymal Stem Cells. *Front Neurol*. 2020;10:1311. doi:10.3389/fneur.2019.01311

37. Patrícia Maria Dias Correia. Identification and characterization of potential therapeutic targets for spinal cord repair. Published online 2017.

38. Duan H, Ge W, Zhang A, et al. Transcriptome analyses reveal molecular mechanisms underlying functional recovery after spinal cord injury. *Proc Natl Acad Sci*. 2015;112(43):13360-13365. doi:10.1073/pnas.1510176112

39. Munro KM, Perreau VM, Turnley AM. Differential Gene Expression in the EphA4 Knockout Spinal Cord and Analysis of the Inflammatory Response Following Spinal Cord Injury. Langmann T, ed. *PLoS ONE*. 2012;7(5):e37635. doi:10.1371/journal.pone.0037635

40. Satzer D, Miller C, Maxon J, et al. T cell deficiency in spinal cord injury: altered locomotor recovery and whole-genome transcriptional analysis. *BMC Neurosci*. 2015;16(1):74. doi:10.1186/s12868-015-0212-0

41. Kantardzhieva A, Alexeeva S, Versteeg I, Wijnholds J. MPP3 is recruited to the MPP5 protein scaffold at the retinal outer limiting membrane. *FEBS J*. 2006;273(6):1152-1165. doi:10.1111/j.1742-4658.2006.05140.x

42. Liu J, Li P, Wang R, et al. High expression of DLG3 is associated with decreased survival from breast cancer. *Clin Exp Pharmacol Physiol*. 2019;46(10):937-943. doi:10.1111/1440-1681.13132

43. Otomo A, Ueda MT, Fujie T, et al. Efficient differentiation and polarization of primary cultured neurons on poly (lactic acid) scaffolds with microgrooved structures. *Sci Rep*. 2020;10(1):6716. doi:10.1038/s41598-020-63537-z

44. Liu J, Han D, Wang Z, et al. Clinical analysis of the treatment of spinal cord injury with umbilical cord mesenchymal stem cells. *Cytotherapy*. 2013;15(2):185-191. doi:10.1016/j.jcyt.2012.09.005

45. Gaete M, Muñoz R, Sánchez N, et al. Spinal cord regeneration in Xenopus tadpoles proceeds through activation of Sox2-positive cells. *Neural Develop*. 2012;7(1):13. doi:10.1186/1749-8104-7-13

46. Amigo J, Opazo J, Jorquera R, et al. The Reprimo Gene Family: A Novel Gene Lineage in Gastric Cancer with Tumor Suppressive Properties. *Int J Mol Sci*. 2018;19(7):1862. doi:10.3390/ijms19071862

47. Saavedra K, Valbuena J, Olivares W, et al. Loss of Expression of Reprimo, a p53-induced Cell Cycle Arrest Gene, Correlates with Invasive Stage of Tumor Progression and p73 Expression in Gastric Cancer. Soutto M, ed. *PLOS ONE*. 2015;10(5):e0125834. doi:10.1371/journal.pone.0125834

48. Stanic K, Quiroz A, Lemus CG, et al. Expression of RPRM/rprm in the Olfactory System of Embryonic Zebrafish (Danio rerio). *Front Neuroanat*. 2018;12:23. doi:10.3389/fnana.2018.00023

49. Lai J, Wang H, Luo Q, et al. The relationship between DNA methylation and Reprimo gene expression in gastric cancer cells. *Oncotarget*. 2017;8(65):108610-108623. doi:10.18632/oncotarget.21296

50. Taylor WR, Stark GR. Regulation of the G2/M transition by p53. *Oncogene*. 2001;20(15):1803-1815. doi:10.1038/sj.onc.1204252

51. Ohki R, Nemoto J, Murasawa H, et al. Reprimo, a New Candidate Mediator of the p53-mediated Cell Cycle Arrest at the G2 Phase. *J Biol Chem*. 2000;275(30):22627-22630. doi:10.1074/jbc.C000235200

52. Kovalevich J, Langford D. Considerations for the Use of SH-SY5Y Neuroblastoma Cells in Neurobiology. In: Amini S, White MK, eds. *Neuronal Cell Culture*. Vol 1078. Methods in Molecular Biology. Humana Press; 2013:9-21. doi:10.1007/978-1-62703-640-5_2

53. Xicoy H, Wieringa B, Martens GJM. The SH-SY5Y cell line in Parkinson's disease research: a systematic review. *Mol Neurodegener*. 2017;12(1):10. doi:10.1186/s13024-017-0149-0

54. Encinas M, Iglesias M, Liu Y, et al. Sequential Treatment of SH-SY5Y Cells with Retinoic Acid and Brain-Derived Neurotrophic Factor Gives Rise to Fully Differentiated, Neurotrophic Factor-Dependent, Human Neuron-Like Cells. *J Neurochem*. 2002;75(3):991-1003. doi:10.1046/j.1471-4159.2000.0750991.x

55. Bae DH, Marino M, Iaffaldano B, et al. Design and Testing of Vector-Producing HEK293T Cells Bearing a Genomic Deletion of the SV40 T Antigen Coding Region. *Mol Ther - Methods Clin Dev*. 2020;18:631-638. doi:10.1016/j.omtm.2020.07.006

56. Merten O-W, Hebben M, Bovolenta C. Production of lentiviral vectors. *Mol Ther*. Published online 2016:14.

57. da Cruz e Silva OAB, Vieira SI, Rebelo S, da Cruz e Silva EF. A Model System to Study Intracellular Trafficking and Processing of the Alzheimer's Amyloid Precursor Protein. *Neurodegener Dis*. 2004;1(4-5):196-204. doi:10.1159/000080986

58. Yeo NC, Chavez A, Lance-Byrne A, et al. An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat Methods*. 2018;15(8):611-616. doi:10.1038/s41592-018-0048-5

59. van Groningen T, Akogul N, Westerhout EM, et al. A NOTCH feed-forward loop drives reprogramming from adrenergic to mesenchymal state in neuroblastoma. *Nat Commun*. 2019;10(1):1530. doi:10.1038/s41467-019-09470-w

60. Zhang J, Niu H, Zhao ZJ, et al. CRISPR/Cas9 Knockout of Bak Mediates Bax Translocation to Mitochondria in response to TNF *α* /CHX-induced Apoptosis. *BioMed Res Int*. 2019;2019:1-8. doi:10.1155/2019/9071297

61. Prasuhn J, Mårtensson CU, Krajka V, Klein C, Rakovic A. Genome-Edited, TH-expressing Neuroblastoma Cells as a Disease Model for Dopamine-Related Disorders: A Proof-of-Concept Study on DJ-1-deficient Parkinsonism. *Front Cell Neurosci*. 2018;11:426. doi:10.3389/fncel.2017.00426

62. Chen Y, Yang Z, Wei L, et al. Yes-associated protein protects and rescues SH-SY5Y cells from ketamine-induced apoptosis. *Mol Med Rep*. 2020;22(3):2342-2350. doi:10.3892/mmr.2020.11328

63. Wang Z, Lei H, Sun Q. MicroRNA-141 and its associated gene FUS modulate proliferation, migration and cisplatin chemosensitivity in neuroblastoma cell lines. *Oncol Rep*. 2016;35(5):2943-2951. doi:10.3892/or.2016.4640

64. Berger A, Lang R, Moritz K, et al. Galanin Receptor Subtype GalR2 Mediates Apoptosis in SH-SY5Y Neuroblastoma Cells. *Endocrinology*. 2004;145(2):500-507. doi:10.1210/en.2003-0649

65. Khatib T, Whiting A, Chisholm DR, Redfern C, Müller B, McCaffery P. A Bioluminescence Reporter Assay for Retinoic Acid Control of Translation of the GluR1 Subunit of the AMPA Glutamate Receptor. *Mol Neurobiol*. 2019;56(10):7074-7084. doi:10.1007/s12035-019-1571-9

66. Pu J, Mao Y, Xu L, Zheng T, Zhang B. Stable cell lines of human SH-SY5Y uniformly expressing wild-type or mutant-type FERM domain containing 7 gene. *Exp Ther Med*. 2017;14(3):2277-2283. doi:10.3892/etm.2017.4730

67. Fan W, Chang S, Shan X, et al. Transcriptional profile of SH-SY5Y human neuroblastoma cells transfected by Toxoplasma rhoptry protein 16. *Mol Med Rep*. 2016;14(5):4099-4108. doi:10.3892/mmr.2016.5758

68. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613. doi:10.1093/nar/gky1131

69. Charlesworth G, Plagnol V, Holmström KM, et al. Mutations in ANO3 Cause Dominant Craniocervical Dystonia: Ion Channel Implicated in Pathogenesis. *Am J Hum Genet*. 2012;91(6):1041-1050. doi:10.1016/j.ajhg.2012.10.024

70. Stamelou M, Charlesworth G, Cordivari C, et al. The phenotypic spectrum of DYT24 due to ANO3 mutations: The phenotypic spectrum of DYT24. *Mov Disord*. 2014;29(7):928-934. doi:10.1002/mds.25802

71. DOOFNL Consortium, Smits JJ, Oostrik J, et al. De novo and inherited loss-of-function variants of ATP2B2 are associated with rapidly progressive hearing impairment. *Hum Genet*. 2019;138(1):61-72. doi:10.1007/s00439-018-1965-1

72. Yang W, Liu J, Zheng F, et al. The Evidence for Association of ATP2B2 Polymorphisms with Autism in Chinese Han Population. Zhang XY, ed. *PLoS ONE*. 2013;8(4):e61021. doi:10.1371/journal.pone.0061021

73. Chen M, Laursen SH, Habekost M, et al. Central and Peripheral Nervous System Progenitors Derived from Human Pluripotent Stem Cells Reveal a Unique Temporal and Cell-Type Specific Expression of PMCAs. *Front Cell Dev Biol*. 2018;6:5. doi:10.3389/fcell.2018.00005

74. Park KA, Fehrenbacher JC, Thompson EL, Duarte DB, Hingtgen CM, Vasko MR. Signaling pathways that mediate nerve growth factor-induced increase in expression and release of calcitonin gene-related peptide from sensory neurons. *Neuroscience*. 2010;171(3):910-923. doi:10.1016/j.neuroscience.2010.09.027

75. Sun R-Q, Lawand NB, Lin Q, Willis WD. Role of Calcitonin Gene-Related Peptide in the Sensitization of Dorsal

Horn Neurons to Mechanical Stimulation After Intradermal Injection of Capsaicin. *J Neurophysiol*. 2004;92(1):320-326. doi:10.1152/jn.00086.2004

76. Sekiguchi T. The Calcitonin/Calcitonin Gene-Related Peptide Family in Invertebrate Deuterostomes. *Front Endocrinol*. 2018;9:695. doi:10.3389/fendo.2018.00695

77. Su Y-C, Feng Y-H, Wu H-T, et al. Elovl6 is a negative clinical predictor for liver cancer and knockdown of Elovl6 reduces murine liver cancer progression. *Sci Rep*. 2018;8(1):6586. doi:10.1038/s41598-018-24633-3

78. Nagarajan R, Le N, Mahoney H, Araki T, Milbrandt J. Deciphering peripheral nerve myelination by using Schwann cell expression profiling. *Proc Natl Acad Sci*. 2002;99(13):8998-9003. doi:10.1073/pnas.132080999

79. Lim J, Choi H-S, Choi HJ. Estrogen-related receptor gamma regulates dopaminergic neuronal phenotype by activating GSK3β/NFAT signaling in SH-SY5Y cells. *J Neurochem*. 2015;133(4):544-557. doi:10.1111/jnc.13085

80. Hua T, Wang X, Chi S, et al. Estrogen-related receptor γ promotes the migration and metastasis of endometrial cancer cells by targeting S100A4. *Oncol Rep*. Published online 2018:10.

81. Sivera R, Frasquet M, Lupo V, et al. Distribution and genotype-phenotype correlation of GDAP1 mutations in Spain. *Sci Rep*. 2017;7(1):6677. doi:10.1038/s41598-017-06894-6

82. Googins MR, Woghiren-Afegbua AO, Calderon M, St. Croix CM, Kiselyov KI, VanDemark AP. Structural and functional divergence of GDAP1 from the glutathione S-transferase superfamily. *FASEB J*. 2020;34(5):7192-7207. doi:10.1096/fj.202000110R

83. Maliekal P, Sokolova T, Vertommen D, Veiga-da-Cunha M, Van Schaftingen E. Molecular Identification of Mammalian Phosphopentomutase and Glucose-1,6-bisphosphate Synthase, Two Members of the α-D-Phosphohexomutase Family. *J Biol Chem*. 2007;282(44):31844-31851. doi:10.1074/jbc.M706818200

84. von Mollard GF, Stah B, Khokhlatchev A, Siidhofl ThomasC, Jahn R. Rab3C Is a Synaptic Vesicle Protein That Dissociates from Synaptic Vesicles after Stimulation of Exocytosis. *J Biol Chem*. 1994;269(15):10971-10974.

85. Chang Y-C, Su C-Y, Chen M-H, Chen W-S, Chen C-L, Hsiao M. Secretory RAB GTPase 3C modulates IL6-STAT3 pathway to promote colon cancer metastasis and is associated with poor prognosis. *Mol Cancer*. 2017;16(1):135. doi:10.1186/s12943-017-0687-7

86. Bon E, Driffort V, Gradek F, et al. SCN4B acts as a metastasis-suppressor gene preventing hyperactivation of cell migration in breast cancer. *Nat Commun*. 2016;7(1):13648. doi:10.1038/ncomms13648

87. Ji X, Saha S, Gao G, et al. The Sodium Channel β4 Auxiliary Subunit Selectively Controls Long-Term Depression in Core Nucleus Accumbens Medium Spiny Neurons. *Front Cell Neurosci*. 2017;11. doi:10.3389/fncel.2017.00017

88. Reardon BJ, Hansen JG, Crystal RG, et al. Vitamin D-responsive SGPP2 variants associated with lung cell expression and lung function. *BMC Med Genet*. 2013;14(1):122. doi:10.1186/1471-2350-14-122

89. Hill RZ, Hoffman BU, Morita T, et al. The signaling lipid sphingosine 1-phosphate regulates mechanical pain. *eLife*. 2018;7:e33285. doi:10.7554/eLife.33285

90. Beaubien F, Raja R, Kennedy TE, Fournier AE, Cloutier J-F. Slitrk1 is localized to excitatory synapses and promotes their development. *Sci Rep*. 2016;6(1):27343. doi:10.1038/srep27343

91. Xu M, Knox AJ, Michaelis KA, et al. Reprimo (RPRM) Is a Novel Tumor Suppressor in Pituitary Tumors and Regulates Survival, Proliferation, and Tumorigenicity. *Endocrinology*. 2012;153(7):2963-2973. doi:10.1210/en.2011-2021

92. da Rocha JF, da Cruz e Silva OAB, Vieira SI. Analysis of the amyloid precursor protein role in neuritogenesis reveals a biphasic SH-SY5Y neuronal cell differentiation model. *J Neurochem*. 2015;134(2):288-301. doi:10.1111/jnc.13133

93. Zeng H, Yu B, Liu N, et al. Transcriptomic analysis of α-synuclein knockdown after T3 spinal cord injury in rats. *BMC Genomics*. 2019;20(1):851. doi:10.1186/s12864-019-6244-6

94. Altimimi HF, Schnetkamp PPM. Na+-dependent Inactivation of the Retinal Cone/Brain Na+/Ca2+-K+ Exchanger NCKX2. *J Biol Chem*. 2006;282(6):3720-3729. doi:10.1074/jbc.M609285200

95. Tang P, Zhang Y, Chen C, et al. In Vivo Two-Photon Imaging of Axonal Dieback, Blood Flow and Calcium Influx with Methylprednisolone Therapy after Spinal Cord Injury. *Sci Rep*. 2015;5(1):9691. doi:10.1038/srep09691

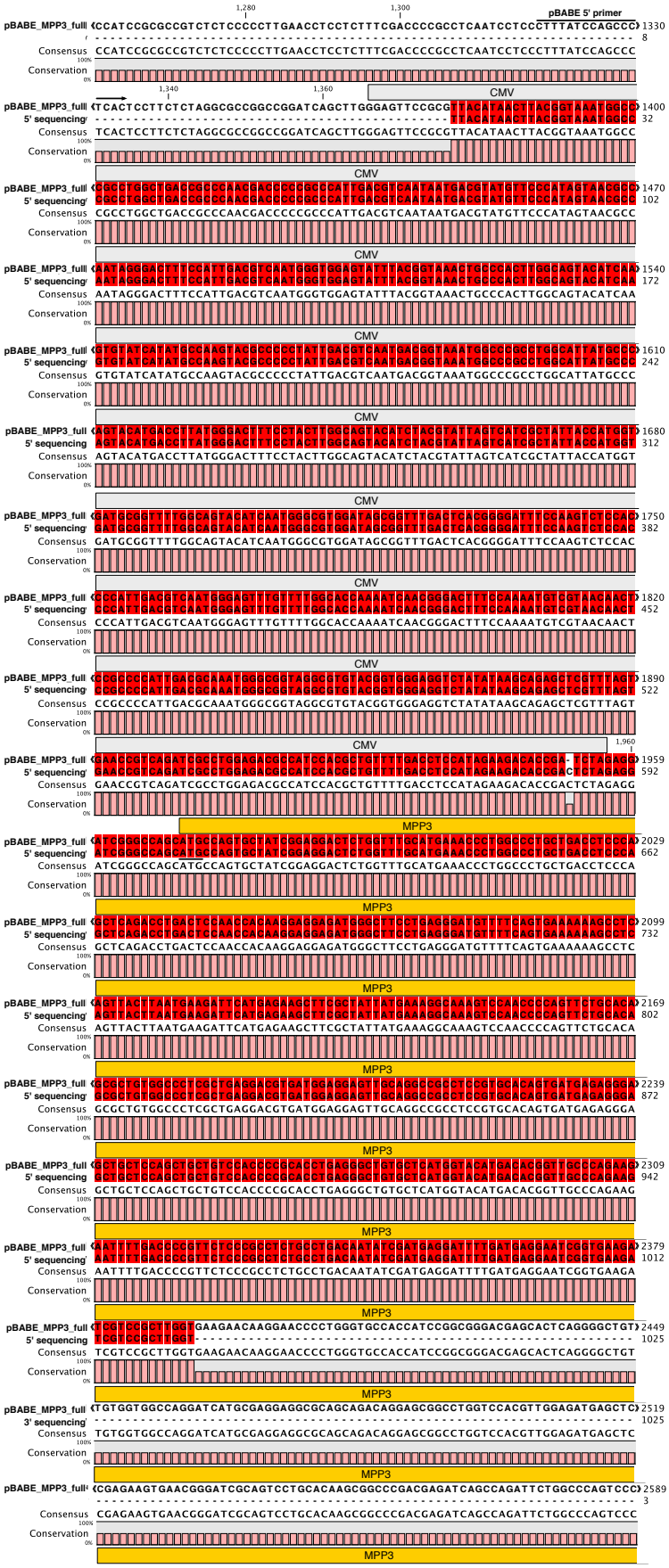
96. Assas BM, Pennock JI, Miyan JA. Calcitonin gene-related peptide is a key neurotransmitter in the neuro-immune axis. *Front Neurosci*. 2014;8. doi:10.3389/fnins.2014.00023

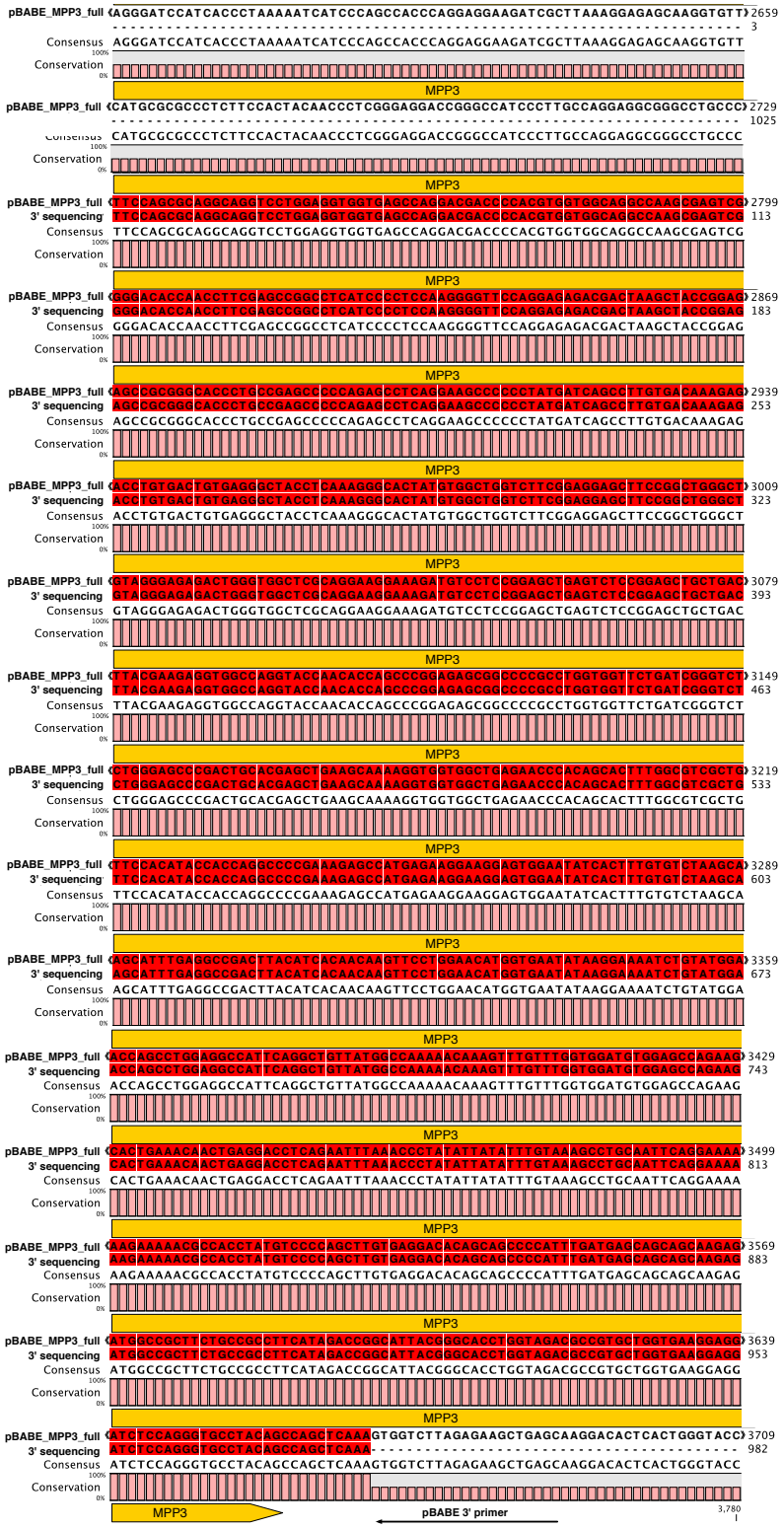
97. Vogt Weisenhorn DM, Roback LJ, Kwon JH, Wainer BH. Coupling of cAMP/PKA and MAPK Signaling in Neuronal Cells Is Dependent on Developmental Stage. *Exp Neurol*. 2001;169(1):44-55. doi:10.1006/exnr.2001.7651

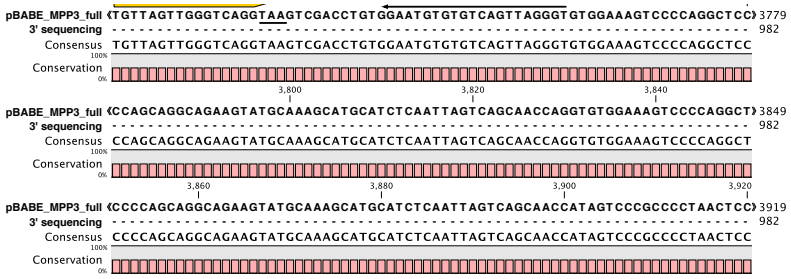
98. Tran AP, Warren PM, Silver J. The Biology of Regeneration Failure and Success After Spinal Cord Injury. *Physiol Rev*. 2018;98(2):881-917. doi:10.1152/physrev.00017.2017

99. Scholpa NE, Schnellmann RG. Mitochondrial-Based Therapeutics for the Treatment of Spinal Cord Injury: Mitochondrial Biogenesis as a Potential Pharmacological Target. *J Pharmacol Exp Ther*. 2017;363(3):303-313. doi:10.1124/jpet.117.244806

6. Annex


pBABE 5' primer
CMV
MPP3
pBABE_MPP3_full
5' sequencing
3' sequencing
Consensus
Conservation


pBABE_MPP3_full
Consensus
Conservation
MPP3
3' sequencing
pBABE 3' primer



Supplementary Figure 1. Alignment between the complete sequencing readout of the selected positive clone DNA, and the pBABE fl-MPP3 and pBABE t-MPP3 sequences using 5' Check_MPP3_full forward and pBABE 3'_reverse primers.

Supplementary Table 1. Studies on potential therapies for SCI that include a transcriptomics analysis.

Study	Authors	Year	Species	Treatment	Analysis	Main conclusion
Transcriptome analyses reveal molecular mechanisms underlying functional recovery after spinal cord injury	Duan H, Ge W, Zhang A, Xi Y, Chen Z, Luo D, Cheng Y, Fan KS, Horvath S, Sofroniew MV, Cheng L, Yang Z, Sun YE, Li X	2015	Rattus norvegicus	NT3 treatment with chitosan tubular scaffold	Affymetrix GeneChip (Rat Genome 230) 2.0	NT3-chitosan enhances new neurogenesis and angiogenesis, and reduces inflammatory responses
Differential gene expression in the EphA4 knockout spinal cord and analysis of the inflammatory response following spinal cord injury	Munro KM, Perreau VM, Turnley AM	2012	Mus musculus	EphA4 knockout rats	Affymetrix Mouse All-Exon ST Array GeneChipsTM	Subtle alterations in the neuroinflammatory response in injured EphA4 knockout spinal cords may contribute to the regeneration and recovery following SCI
T cell deficiency in spinal cord injury: altered locomotor recovery and whole-genome transcriptional analysis	Satzer D, Miller C, Maxon J, Voth J, DiBartolomeo C, Mahoney R, Dutton JR, Low WC, Parr AM	2015	Rattus norvegicus	T-cell deficient athymic nude rats	HiSeq 2000 (Illumina)	AN rats have delayed tissue death and limit long-term recovery. T cell inhibition combined with other neuroprotective treatment may thus be a promising therapeutic avenue
Transcriptomic analysis of α-synuclein knockdown after T3 spinal cord injury in rats	Zeng H, Yu B, Yang Y, Xing H, Liu X, Zhou M	2019	Rattus norvegicus	Lentiviral vector-mediated knockdown of α-Syn	HiSeq xten (Illumina)	Knockdown of α-Syn after SCI enhance motor function and promote neurogenesis probably through enhancing cholinergic signaling pathways and neuroreceptor interactions
Cell cycle and complement inhibitors may be specific for treatment of spinal cord injury in aged and young mice: transcriptomic analyses	Takano M, Kawabata S, Shibata S, Yasuda A, Nori S, Tsuji O, Nagoshi N, Iwanami A, Ebise H, Horiuchi K, Okano H, Nakamura M	2017	Mus musculus	Neural stem cell (NSC) transplantation	Affymetrix Mouse Genome 430 2.0 Array	Aged mice with SCI exhibit enhanced regenerative capability following NSC transplantation, even though they suffer more severely from the SCI than do young mice
A transcriptomic study of probenecid on injured spinal cords in mice	Zhang Y, Wang S, Chen J, Hu J, Lü H	2020	Mus musculus	Probenecid	HiSeq (Illumina)	Probenecid treatment following SCI can play an anti-inflammatory role by inhibiting the infiltration of inflammatory cells

Transcriptome Analysis Reveals the Effects of Troxerutin and Cerebroprotein Hydrolysate Injection on Injured Spinal Cords in Rats	Wang L, Wang H, Tang K, Zhong W, Chen Z, Quan Z	2020	Rattus norvegicus	Troxerutin and Cerebroprotein Hydrolysate	Hiseq2000 Truseq SBS Kit v3-HS (Illumina)	TCH injection with the right dose is effective for the recovery of locomotion function and repairing of the damaged tissue in the spinal cord; It also has a role in the regulation of 443 DEGs involved in multiple functions, including locomotion, blood vessel morphogenesis, thiamine metabolism, Hippo signaling pathway, and axon guidance
Effect of VX-765 on the transcriptome profile of mice spinal cords with acute injury	Chen J, Chen Y, Wang S, Duan F, Shi Y, Ding S, Hu J, LÜ H	2020	Mus musculus	VX-765	Hiseq (Illumina)	Application of VX-765 in the acute phase can improve the local microenvironment of SCI by inhibiting caspase-1
DNA microarray analysis of the contused spinal cord: Effect of NMDA receptor inhibition	Nesic O, Svrakic NM, Xu G-Y, McAdoo D, Westlund KN, Hulsebosch CE, Zeiming Ye, Galante A, Soteropoulos P, Tolias P, Young W, Hart RO, Perez-Polo JR	2002	Rattus norvegicus	NMDA receptor inhibition	Affymetrix U34 microarray	SCI-induced NMDA receptor activation is one of the dominant, early signals after trauma that leads to changes in mRNA levels of a number of genes relevant to recovery processes. NMDA receptor inhibition reverses the effect of SCI on about 50% of the SCI-affected mRNAs
Long-Lasting Sprouting and Gene Expression Changes Induced by the Monoclonal Antibody IN-1 in the Adult Spinal Cord	Bareyre F, Haudenschild B, Schwab M	2002	Rattus norvegicus	mAb IN-1	Affymetrix rat genome U34A GeneChip in Affymetrix fluidics station 400	IN-1 antibody induces endogenous sprouting and enhances the compensative capacity of adult motor CST fibers to reorganize in a target-specific manner
Gene expression profiling in anti-CD11d mAb-treated spinal cord-injured rats	Gris P, Tighe A, Thawer S, Hemphill A, Oatway M, Weaver L, Dekaban G, Brown A	2009	Rattus norvegicus	Acute administration of a mAb CD11d subunit of the leukocyte CD11d/CD18 integrin	Affymetrix GeneChip Scanner 3000	Reduced expression of pro-inflammatory cytokines and increased expression of inflammatory mediators that promote wound healing and the expression of scar proteins predicted to improve nerve growth. These changes in gene expression may reflect changes in the types of macrophages that populate the lesions in anti-CD11d mAb-treated rats
Axotomy-Induced Smad1 Activation Promotes Axonal Growth in Adult Sensory Neurons	Zou H, Ho C, Wong K, Tessier-Lavigne M	2009	Mus musculus (Peripheral lesion!)	Smad1 Activation	Affymetrix U34A chips	Activation of Smad1 by axotomy is a key component of the transcriptional switch that promotes an enhanced growth state of adult sensory neurons