



INTERNATIONAL DOCTORAL
SCHOOL OF THE USC

Sara
Veiga Rúa

PhD Thesis

Deciphering Autism Spectrum
Disorder genomic variation
through the characterization of
zebrafish and human stem cell-
based models

Santiago de Compostela, 2023

Doctoral Programme in Molecular Medicine

DOCTORAL THESIS

**DECIPHERING AUTISM SPECTRUM
DISORDER GENOMIC VARIATION
THROUGH THE
CHARACTERIZATION OF ZEBRAFISH
AND HUMAN STEM CELL-BASED
MODELS**

Author

Sara Veiga Rúa

Supervisors: Laura Elena Sánchez Piñón and Catarina Allegue Toscano

Tutor: Laura Elena Sánchez Piñón

PHD PROGRAMME IN MOLECULAR MEDICINE

SANTIAGO DE COMPOSTELA / LUGO

CONFLICT OF INTEREST STATEMENT

The doctoral candidate, Sara Veiga Rúa DNI 78810873-Q, declares no conflict of interest related to her doctoral thesis.

FIGURES AUTHORSHIP STATEMENT

The doctoral candidate, Sara Veiga Rúa DNI 78810873-Q, declares that all the figures included in the present document have been created by herself, except for Figures 10-15. These figures contain graphs available in public databases under a Creative Commons license.

ANIMAL EXPERIMENTATION STATEMENT

I, Sara Veiga Rúa, with DNI 78810873-Q, declare that this doctoral thesis was conducted within the research project *El pez cebra y su aplicación en biomedicina, acuicultura y medio ambiente* (Exp. 01/20/LU-003), approved by the Animal Experimentation Ethical Committee (CEAA) of the University of Santiago de Compostela (MR110250) (Annexe I, Annexe II), according to the current animal welfare legislation and to the standard protocols (*Real Decreto 53/2013, Directiva 2010-63-EU*). All procedures were performed in the animal facility located in the Zoology, Genetics and Anthropology Department, Veterinary Faculty, Lugo (AE-LU-003), which REGA code of authorization is ES270280346401. All procedures were conducted by Sara Veiga Rúa, who has been qualified for A, B and C animal research functions in fish, amphibians, and rodents (Annexe III).

HUMAN INDUCED PLURIPOTENT STEM CELL EXPERIMENTATION STATEMENT

I, Sara Veiga Rúa, with DNI 78810873-Q, declare that this doctoral thesis was conducted within the research project *Enfoque integral y translacional del diagnóstico y tratamiento de trastornos del neurodesarrollo (TEA, TDAH, DI) y del TOC: Una aproximación genética y funcional*. This project has been reviewed and approved by the Ethical Committee of Santiago-Lugo Research dependent of the *Consellería de Sanidade da Xunta de Galicia* (Ref. 2020/400, Annexe IV). In addition, the *Banco Nacional de Líneas Celulares, Instituto de Salud Carlos III* has reviewed and approved the use of the human induced pluripotent stem cell line Ctrl.BS-PBiPS37-Sv4F-1 in this research project (MTA displayed in Annexe V). All procedures were performed according to current legislation (Ley 14/2007).

FUNDING

This work has been conducted with the support of the *Consellería de Educación, Universidade e Formación Profesional* (ED431C-2018/28, ED431C-2022/33) and the *Instituto Carlos III-Fondos FEDER* (PI19/00809, 2020-2022; PI22/00208, 2023-2025).

During the development of her doctoral thesis, Sara Veiga Rúa's work was funded by several agencies. From 01/10/2018 to 30/05/20219, she was supported by *Contratos de Iniciación a la Investigación 2018* (Universidade de Santiago de Compostela). From 31/05/2019 to 22/09/2019 by *Ayudas de Apoyo a la Etapa Predoctoral* (ED481A-2019/049, Xunta de Galicia). From 23/09/2019 to 31/05/2023, she was supported by *Ayudas para contratos predoctorales FPU 2018* (FPU18/00402, *Ministerio de Educación y Formación Profesional, Gobierno de España*). Also, she received additional funds from *Ayudas complementarias de movilidad destinadas a beneficiarios del programa de Formación del Profesorado Universitario* (EST22/00183), to do a short stay at the University of Lund (Sweden).

AGRADECEMENTOS

Foi toda unha casualidade que me tocasse poñerme a escribir os agradecementos da tese ás portas do Nadal, pero non se me ocorre unha época máis axeitada para pensar en todas as persoas que estiveron ao meu lado neste longo camiño. Podería escribir 300 páxinas máis de agradecementos, pero vou intentar ser breve:

En primeiro lugar, quero comezar agradecéndolles o seu apoio constante as miñas directoras de tese, Cata e Laura. Tiven a inmensa sorte de facer a tese en dous laboratorios nos que sempre me sentín apoiada, escoitada e motivada, e todo iso é grazas a vos.

Quero aproveitar para darlles as grazas a Sara Pagáns e Isaac Canals por acollerme cos brazos abertos nos seus laboratorios durante as miñas estancias predoutorais. Son mellor investigadora grazas a vós e ao equipo que vos rodea.

Tampouco quero esquecerme de mencionar aos nosos colaboradores, Susana Bravo, James Perkins e Sergio Alías, por axudarme a darlle sentido aos datos sen perder a cabeza polo camiño.

Ao equipo do Departamento de Xenética. Aos profesores que me deron a oportunidade de dar docencia. As técnicas de laboratorio (Peque, Lucía, Cris, Susana) por estar sempre dispostas a botar unha man e resolver dúbidas. A Sonia e Mónica pola vosa axuda cos odiados trámites burocráticos. E por suposto, aos investigadores pre e posdoutorais do grupo, grazas por acollerme como unha máis nas miñas escapadas a Lugo. En especial, quero darlle as grazas a Alba. Sabes que non podería escribir uns agradecementos de tese sen terte presente. Grazas por ensinarme todo o que sei dos peixes cebrá, non sei que houbese feito sen ti.

A todo o grupo de Medicina Xenómica, en especial a Ángel pola confianza que sempre demostraches neste proxecto e polo teu apoio incondicional. Ao equipo de secuenciación (Pancho, Quique, Sonia), de cariotipado (Luz, Raquel) e de NGS (Bea, Elena) gracias pola vosa axuda cos experimentos desta tese. A Esther e a Vicente, pola súa axuda inestimable co papeleo, que sempre me fai perder un pouco a paciencia. A Cris, Raquel, Montse, e Eduardo por crear un gran ambiente de traballo no CiMUS. Ao fantástico equipo de Farmacoxenómica, grazas por todos os momentos compartidos no SSL1, en especial a Almudena que me acompañou dende o principio. A Damián, Sara e Silvia, ánimo con esa escritura que xa queda pouco, sabedes que estou aquí para o que precisades. A Jorge, alégrome moito de que por fin poidas defender a tese, es un investigador fantástico e telo máis que merecido. A Raquel e a Fernando, foi un pracer compartir esta etapa de tese con vos, da gusto compartir laboratorio con xente tan simpática e traballadora.

Quero dedicarlles tamén unhas palabras as persoas que máis tempo pasaron comigo, as miñas maravillosas compañeiras de laboratorio, que tamén se converteron en amigas. Xuro que a partir de agora vou deixar que me fagades todas as preguntas que queirades e que irei máis a tomar o café. Elena, es a mellor compañeira de traballo que me podería ter tocado. As comidas sen ti fan os días un pouco máis tristes. Miguel, creo que es a persoa que máis tivo que sufrir os meus cambios de humor durante a escritura desta tese, grazas por non tirarme pola ventá do laboratorio. Non te esquezas de nós en Canadá. Lucía, gracias polo teu optimismo contaxioso, es un ser de luz. Alba S, unha das mellores cousas que me levo desta tese é a túa amizade. Non podería estar máis orgullosa de ti.

As nosas compañeiras de Cardio, gracias por todos os momentos compartidos. María, agora que deposito a tese, espero parecerme aínda máis a ti.

As miñas amigas de fora do laboratorio, non me imaxino como houbese sido enfrontarme á tese sen o voso apoio, son afortunada de tervos na miña vida.

Por último, á miña familia. Álex, María, Pablo, Iván, prometo ir a vervos máis a miúdo agora que a tese remata. A Pura, miña nai, grazas por transmitirme a túa curiosidade polo mundo. A Juan, meu pai, grazas por ensinarme a tirar para adiante pase o que pase. Sen eses valores fundamentais, son houbese sido capaz de sacar adiante este traballo. Álex, despois de MESES preguntándome todas as semanas se xa rematei de escribir, por fin podo dicir que si.

INDICES

ABBREVIATION LIST

µg: microgram	BCL2: BCL2 apoptosis regulator
µL: microliter	BCLAF1: BCL2 associated transcription factor 1
µM: micromolar	bFGF: basic fibroblast growth factor
µS: micro-Siemens	BMP: bone morphogenetic protein
aa: amino acid	bp: base pairs
AAV: Adeno-Associated Virus	BSA: Bovine Serum Albumin
ABA: Applied Behaviour Analysis	BTB/POZ: broad-complex, tramtrack, and bric-à-brac/poxvirus and zinc finger
ABE: Adenine Base Editor	BTBD8: BTB domain containing 8
ACTL6B: actin like 6B	bZIP: basic leucine zipper
ACVR1B: activin A receptor type 1B	CACNA1C: calcium voltage-gated channel subunit alpha1 C
ACVR1C: activin A receptor type 1C	cas: CRISPR-associated
ADCY6: adenylate cyclase 6	CBT: Cognitive-Behavioural Therapy
ADHD: Attention-Deficit/Hyperactivity Disorder	CDLC: Chemically Defined Lipid Concentrate
AFP: alpha fetoprotein	CDM: Cortical Differentiation Medium
AGD: Autism Genetic Database	cDNA: complementary DNA
AHNAK: AHNAK nucleoprotein	CFuPN: corticofugal projection neurons
Akt: protein kinase B	chr: chromosome
AMPA: α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid	cm: centimetres
ANKRD11: ankyrin repeat domain containing 11	CNS: Central Nervous System
ANOVA: Analysis of Variance	CNTN2: contactin 2
ANTXR1: ANTXR cell adhesion molecule 1	Cntnap2: contactin associated protein 2
AP-1: activator protein 1	CNVs: Copy Number Variation
AP-2: adaptor protein 2	COL1A1: collagen type I alpha 1 chain
AP2A1: adaptor related protein complex 2 subunit alpha 1	COL1A2: collagen type I alpha 2 chain
AP2B1: adaptor related protein complex 2 subunit beta 1	COL3A1: collagen type III alpha 1 chain
AP3B2: adaptor related protein complex 3 subunit beta 2	COL6A1: collagen type VI alpha 1 chain
APA: American Psychiatric Association	CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats
APS: ammonium persulfate	CRISPRa: CRISPR activation
aRG: apical radial glia	CRISPRi: CRISPR interference
ASD: Autism Spectrum Disorders	crRNA: CRISPR RNA
ATF: activating transcription factor	CTIP2/BCL11B: BCL11 transcription factor B
ATP: adenosine triphosphate	CUL3: cullin 3
BBB: Blood-Brain Barrier	DABCO: 1,4-diazabicyclo[2.2.2]octane
	dCas9: dead Cas9

DCTN1: dynactin subunit 1	GAD67: glutamic acid decarboxylase 67
DCX: doublecortin	GAPDH: glyceraldehyde-3-phosphate dehydrogenase
ddCt: double delta cycle threshold	GCN1: GCN1 activator of EIF2AK4
DEPC: diethyl pyrocarbonate	GEMs: Gel Bead-In-EMulsions
DLX2: distal-less homeobox 2	GFAP: glial fibrillary acidic protein
DLX2: distal-less homeobox 2	GFP: Green Fluorescent Protein
DLX5: distal-less homeobox 5	GMEM: Glasgow's MEM
DMEM/F12: Dulbecco's Modified Essential Medium and Ham's F-12 Medium	GO: gene ontology
DNA: Deoxyribonucleic acid	GPCR: G protein-coupled receptor
DNase: Deoxyribonuclease	GWAS: Genome Wide Association Studies
DNM1: dynamin 1	h: hours
DNM2: dynamin 2	H₂O₂: hydrogen peroxide
dNTPs: deoxynucleotide triphosphates	HCl: hydrochloric acid
dpf: days post-fecundation	HDR: Homology-Directed Repair
DSBs: Double-Strand Breaks	HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5 th edition	hESCs: human embryonic stem cells
DTT: dithiothreitol	HGNC: HUGO Gene Nomenclature Database
EBF1: EBF transcription factor 1	hiPSCs: human induced pluripotent stem cells
EBSS: Earle's Balanced Salt Solution	HPA: hypothalamus-pituitary-adrenal axis
ECM: extracellular matrix	hpf: hours post-fecundation
EDTA: Ethylenediaminetetraacetic acid	HPI: hypothalamic-pituitary-interrenal axis
EGFLAM: EGF like, fibronectin type III and laminin G domains	HT: heterozygous
EMA: European Medical Administration	ICC: immunocytochemistry
EMX1: empty spiracles homeobox 1	ICD: International Classification of Diseases
ENU: N-ethyl-N-nitrosourea	ICE: Interference of CRISPR Edits
EOMES: comesodermin	IL-17: interleukin 17
ESC: Embryonic Stem Cells	indel: insertion or deletion
EYS: eyes shut homolog	iPSCs: induced Pluripotent Stem Cell
FAM: fluorescein	ITS-A: Insulin-Transferrin-Selenium-Sodium Pyruvate
FAM107A: family with sequence similarity 107 member A	JUN: Jun proto-oncogene, AP-1 transcription factor subunit
FBS: Foetal Bovine Serum	JUNB: JunB proto-oncogene, AP-1 transcription factor subunit
FDA: United States Food and Drug Administration	JUND: JunD proto-oncogene, AP-1 transcription factor subunit
FDR: False Discovery Rate	kb: kilobases
FOS: Fos proto-oncogene, AP-1 transcription factor subunit	KCl: Potassium Chloride
FOSB: FosB proto-oncogene, AP-1 transcription factor subunit	KCNH1: potassium voltage-gated channel subfamily H member 1
FOSL1: FosB proto-oncogene, AP-1 transcription factor subunit	Klf4: KLF transcription factor 4
FOSL2/Fosl2/fosl2: FOS like 2, AP-1 transcription factor subunit (human/mouse/zebrafish)	KO: knockout
GABA: gamma-aminobutyric acid	

- KOH:** Potassium hydroxide
KOSR: Knockout Serum Replacement
KPa: kilo Pascal
LATS1/2: large tumor suppressor kinase 1/2
LB: Luria-Bertani
LC: liquid chromatography
Lin28: Lin-28 homolog A
LINC01716: long intergenic non-protein coding RNA 1716
LINC02929: long intergenic non-protein coding RNA 2929
lncRNA: long noncoding RNA
M: molar (mol/m³, mol/L)
MAF: MAF bZIP transcription factor
MAF: minor allele frequency
MALDI: matrix-assisted laser desorption ionization
MAP2: microtubule associated protein 2
MAP2K: mitogen-activated protein kinase kinase
MAP3K: mitogen-activated protein kinase kinase kinase
MAPK: mitogen-activated protein kinase
MAPK8: mitogen-activated protein kinase 8
MAPT: microtubule associated protein tau
MARK2: microtubule affinity regulating kinase 2
MEA: Multi-Electrode Arrays
MEFs: Murine Embryonic Fibroblasts
MEM: Minimum Essential Medium
mg: milligram
MgCl₂: magnesium chloride
MIM: Mendelian Inheritance in Man
min: minute
miRNA: micro RNA
mL: millilitre
mM: millimolar
MO: morpholino
mRNA: messenger RNA
MS: mass spectrometry
MTT: 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide
MYC: MYC proto-oncogene, bHLH transcription factor
MYCNUT: MYCN upstream transcript
MYRIP: myosin VIIA and Rab interacting protein
NANOG: Nanog homeobox
NaOH: Sodium hydroxide
NaSSA: Noradrenergic and specific serotonergic antidepressant
nCas9: Cas9 nickase
NCBI: National Center for Biotechnology Information
NDS: normal donkey serum
NES: nestin
ng: nanogram
NGS: next generation sequencing.
NHEJ: non-homologous end joining
nL: nanoliter
nm: nanometres
NMM: Neural Maintenance Medium
NRP1: neuropilin 1
NRP2: neuropilin 2
NSCs: Neural Stem Cells
nTPM: normalized transcript per million
°C: degrees celsius
OCT: Optimal cutting temperature compound
OR51B6: olfactory receptor family 51 subfamily B member 6
OSKM: Pou5f1, Sox2, Klf4 and Myc
OSNL: Pou5f1, Sox2, Nanog, and Lin28
p38: mitogen-activated protein kinase p38
p53: cellular tumor antigen p53
PAM: Protospacer Adjacent Motif
PAX6: paired box 6
PBMCs: Peripheral Blood Mononuclear Cells
PBS: Phosphate-Buffered Saline
PCR: Polymerase Chain Reaction
PDGFRA: platelet derived growth factor receptor alpha
PECS: Picture Exchange Communication System
PEI: polyethylenimine
PFA: paraformaldehyde
PHACTR3: phosphatase and actin regulator 3
PMSF: phenylmethylsulfonyl fluoride
POU5F1: POU class 5 homeobox 1
PZMs: Post-Zygotic Mutations
R²: coefficient of determination

RALGPS1: Ral GEF with PH domain and SH3 binding motif 1
RI: rock inhibitor (Y-27632)
RIPA: radioimmunoprecipitation buffer
RNA: ribonucleic acid
RNA-seq: RNA sequencing
rNTP: ribonucleotide triphosphates
ROCK1: Rho associated coiled-coil containing protein kinase 1
RPL19: ribosomal protein L19
rpm: revolutions per minute
RT: reverse transcriptase
RT-qPCR: quantitative reverse transcription PCR
s: second
SATB2: SATB homeobox 2
SBT: Social-Behaviour Therapy
scRNA-seq: single cell RNA sequencing
SD: standard deviation
SDS: Sodium Dodecyl Sulfate
SFARI: Simons Foundation Autism Research Initiative
sgRNA: single guide RNA
SGTA: small glutamine rich tetratricopeptide repeat co-chaperone alpha
shRNA: short hairpin RNA
SMARCA2: SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2
SNAP91: synaptosome associated protein 91
SNPs: Single-Nucleotide Polymorphism
SNRIs: Serotonin-Norepinephrine Reuptake Inhibitor
SOX2: SRY-box transcription factor 2
SpCas9: Streptococcus pyogenes Cas9
SSRIs: Selective Serotonin Reuptake Inhibitors
STK24: serine/threonine kinase 24
STX7: syntaxin 7
STX8: syntaxin 8
STXBP1: syntaxin binding protein 1
SUFU: SUFU negative regulator of hedgehog signalling
SWH: Salvador–Warts–Hippo pathway

SYP: synaptophysin
SYT1: synaptotagmin 1
ta: annealing temperature
TAE: Tris-acetate-EDTA
TAFAZZIN: tafazzin, phospholipid-lysophospholipid transacylase
TALENs: Transcription Activator Like Effector Nucleases
TALs: Transcription Activator-Like Effectors
TAOK1: Thousand And One Kinase 1
TAOK2: Thousand And One Kinase 2
TAOK3: Thousand And One Kinase 3
tau: microtubule-associated protein tau
TBE: Tris-borate-EDTA
TCF7: transcription factor 7
TEMED: Tetramethylethylenediamine
TGFBR1: transforming growth factor beta receptor 1
TGF-β: transforming growth factor beta
TILLING: Targeted Induced Local Lesions in Genomes
TOPO cloning: topoisomerase-based cloning
TPA: 12-O-tetradecanoylphorbol-13-acetate
TPBS: triton-X-100 in PBS
TRA-1-60: T cell receptor alpha locus
TRA-1-81: podocalyxin.
tracrRNA: Trans-activating CRISPR RNA
TSHZ2: teashirt zinc finger homeobox 2
TUBB3: tubulin beta 3 class III
U: units
UMAP: Uniform Manifold Approximation and Projection
UNC5D: unc-5 netrin receptor D
V: volts
WES: Whole Exome Sequencing
WGS: Whole Genome Sequencing
WHO: World Health Organization
WT: wild type
x g: times gravity
YAP1: Transcriptional coactivator YAP1
ZFNs: Zinc Finger Nucleases

GENERAL INDEX

1. INTRODUCTION	33
1.1. NEURODEVELOPMENTAL DISORDERS.....	33
1.2. AUTISM SPECTRUM DISORDERS.	33
1.2.1. Prevalence of ASD.	33
1.2.2. Aetiology of ASD.	34
1.2.2.1. Environmental risk factors.....	34
1.2.2.2. Genetic risk factors.	35
1.2.2.2.1. Technologies applied for the identification of ASD's genetic risk factors.....	35
1.2.2.2.2. Databases of ASD genetic risk factors.....	36
1.2.3. Molecular pathways altered in ASD.....	37
1.2.4. Diagnosis of ASD.	38
1.2.5. Treatment of ASD.....	39
1.3. GENOME EDITING TOOLS IN THE STUDY OF ASD'S GENETIC VARIATION.....	40
1.3.1. The origins of genome editing tools: from meganucleases to ZFNs and TALENs.....	41
1.3.2. CRISPR-Cas9: from prokaryotic immune systems to the new age of genome editing.	42
1.3.2.1. CRISPR-based systems in biomedicine.....	43
1.4. STEM CELLS AS AN <i>IN VITRO</i> MODEL	44
1.4.1. Types of stem cells.	44
1.4.2. hiPSCs as a tool for modelling human diseases and disorders.....	45
1.4.2.1. Reprogramming of somatic cells into iPSCs.	45
1.4.2.1.1. Characterization of hiPSCs.	46
1.4.2.2. Application of genome editing technologies to disease modelling in hiPSCs.	46
1.4.2.3. Human iPSCs in the study of neurodevelopmental disorders.	47
1.4.2.3.1. 2D differentiation of hiPSCs to ASD-related cell types.....	47
1.4.2.3.2. 3D differentiation of hiPSCs for the study of ASD.....	47
1.5. <i>Danio rerio</i> AS AN <i>IN VIVO</i> MODEL FOR NEUROSCIENCE.	48
1.5.1. Anatomical and ecological features of <i>Danio rerio</i>	48
1.5.2. <i>Danio rerio</i> in biomedical research.	49
1.5.2.1. <i>Danio rerio</i> in the study of neurodevelopmental disorders.....	49
1.5.3. Generation of zebrafish models of interest in biomedical research.....	50
1.5.3.1. Transient silencing of genes with morpholinos.	50
1.5.3.2. TILLING.....	51
1.5.3.3. Advanced gene targeting in zebrafish: ZFNs, TALENs and CRISPR-Cas9.	51
1.6. FUNCTIONAL CHARACTERIZATION OF <i>IN VITRO</i> AND <i>IN VIVO</i> ASD MODELS.	52
1.6.1. Characterization of neuronal differentiation, morphology, and function on <i>in vitro</i> models.	52
1.6.2. -OMICS technologies for the study of neurodevelopmental disorders.....	53
1.6.2.1. Application of single-cell RNA sequencing to ASD research.	53
1.6.2.2. Application of proteomic analysis to ASD research.	54
1.6.3. Phenotypic characterization of zebrafish models.	55
1.6.3.1. Morphometric characterization.....	55
1.6.3.2. Behavioural characterization.....	55
1.7. APPLICATION OF <i>IN VITRO</i> AND <i>IN VIVO</i> MODELS OF ASD: FROM THE MOLECULAR MECHANISMS TO PERSONALIZED MEDICINE.	56
1.8. SELECTION OF GENES TO BE STUDIED IN THIS THESIS.....	56

1.8.1. <i>BTBD8</i>	56
1.8.1.1. <i>BTBD8</i> role in synaptic function.	58
1.8.1.2. Phenotypes, diseases, and traits associated with <i>BTBD8</i> variants.	58
1.8.1.3. Orthologous genes of <i>BTBD8</i> in zebrafish.	59
1.8.2. <i>FOSL2</i>	59
1.8.2.1. AP-1 transcriptional complex: composition and function.	60
1.8.2.2. Phenotypes, diseases, and traits associated with <i>FOSL2</i> variants.	61
1.8.2.3. Orthologous genes of <i>FOSL2</i> in zebrafish.	61
1.8.3. <i>TAOK1</i>	61
1.8.3.1. <i>TAOK1</i> role in cellular function.....	63
1.8.3.2. Phenotypes, diseases, and traits associated with <i>TAOK1</i> variants.	64
1.8.3.3. Orthologous genes of <i>TAOK1</i> in zebrafish.	65
1.9. ETHICAL AND LEGAL CONSIDERATIONS REGARDING THE EXPERIMENTAL PROCEDURES OF THIS THESIS.	65
2. JUSTIFICATION AND OBJECTIVES	69
2.1. JUSTIFICATION.....	69
2.2. OBJECTIVES.	69
3. MATERIALS AND METHODS.....	73
3.1. PROTOCOLS USED FOR <i>IN VITRO</i> EXPERIMENTS.	73
3.1.1. <i>Culture of human pluripotent stem cell lines</i>	73
3.1.1.1. Maintenance of hiPSCs.....	73
3.1.1.2. Expansion of hiPSCs.	74
3.1.1.3. Freezing and thawing hiPSCs.....	74
3.1.2. <i>Gene knockout in hiPSCs</i>	75
3.1.2.1. Knockout guide design.	75
3.1.2.2. Cloning guides into plasmid vectors.....	75
3.1.2.2.1. Plasmid purification.....	76
3.1.2.2.2. sgRNA cloning.....	76
3.1.2.2.3. Transformation, purification and plasmid sequencing.....	77
3.1.2.3. Puromycin kill curve for the hiPSCs line Ctrl.BS-PBiPS37-Sv4F-1.....	78
3.1.2.4. Transfection of CRISPR-Cas9 components and antibiotic selection.	78
3.1.3. <i>Guide efficiency analysis</i>	79
3.1.3.1. DNA extraction from hiPSCs.....	79
3.1.3.2. Amplification of the target region for CRISPR-Cas9 in the bulk of transfected cells.	79
3.1.3.3. ICE analysis.....	80
3.1.3.4. Cloning-based efficiency analysis.....	80
3.1.4. <i>Isolation and characterization of edited clones</i>	81
3.1.4.1. Isolation of edited clones.	81
3.1.4.2. Allele isolation and identification.....	82
3.1.4.3. Karyotyping of hiPSCs.....	82
3.1.4.4. Pluripotency analysis by immunocytochemistry.....	82
3.1.4.5. Pluripotency analysis by RT-qPCR.	83
3.1.4.5.1. RNA extraction from hiPSCs.....	83
3.1.4.5.2. cDNA synthesis.....	83
3.1.4.5.3. Desing and efficiency analysis of RT-qPCR primers.	84
3.1.4.5.4. RT-qPCR based on SYBR Green chemistry.....	85
3.1.4.6. Off-target analysis in edited hiPSCs lines.	85
3.1.4.6.1. DNA extraction.....	85
3.1.4.6.2. PCR and Sanger sequencing.....	85
3.1.4.7. Trilineage differentiation.	86
3.1.5. <i>Differentiation of human pluripotent stem cell lines</i>	87
3.1.5.1. Differentiation of hiPSCs to cortical-like neurons.	87
3.1.5.2. Differentiation of hiPSCs to human brain organoids.....	88
3.1.6. <i>Morphometric analysis of neurons derived from hiPSCs</i>	90

3.1.7. Characterization of differentiated cell types using immunofluorescence	91
3.1.7.1. Immunofluorescence of hiPSCs-derived cortical neurons in monolayer.	91
3.1.7.1.1. Fixation of neurons in monolayer.	91
3.1.7.1.2. Immunofluorescence of hiPSCs-derived neurons.	91
3.1.7.2. Immunofluorescence of hiPSCs-derived brain organoids.	92
3.1.7.2.1. Fixation of organoids.	92
3.1.7.2.2. Organoid cryosectioning.	93
3.1.7.2.3. Immunofluorescence of organoids.	93
3.1.8. Proteomic analysis of hiPSCs-derived cortical neurons.	93
3.1.8.1. Protein extraction from hiPSCs-derived cortical neurons.	93
3.1.8.2. Proteomic analysis.	93
3.1.9. Single cell RNA-seq of human brain organoids.	94
3.1.9.1. Dissociation of organoids to single cells.	95
3.1.9.2. Library preparation and sequencing.	96
3.1.9.3. Single-cell RNA-seq data analysis.	98
3.2. PROTOCOLS USED FOR <i>IN VIVO</i> EXPERIMENTS IN ZEBRAFISH.	98
3.2.1. Maintenance of zebrafish.	98
3.2.1.1. Breeding.	98
3.2.1.2. Anaesthesia and euthanasia.	99
3.2.2. Generation of a morphant model of <i>fosl2</i> in zebrafish.	99
3.2.2.1. Microinjection of morpholinos.	99
3.2.2.2. Morphological analysis at 48 hpf.	99
3.2.2.3. Behavioural analysis of the morphants using the light-dark test.	100
3.2.2.4. Alcian blue staining of cartilaginous tissue.	100
3.2.2.5. Knockdown verification by RT-PCR.	101
3.2.2.5.1. RNA extraction from zebrafish embryos.	101
3.2.2.5.2. cDNA synthesis.	101
3.2.2.5.1. RT-PCR.	102
3.2.3. Generation of a mutant zebrafish line using CRISPR-Cas9	102
3.2.3.1. Guide design and synthesis.	102
3.2.3.1.1. In silico knockout guide design.	103
3.2.3.1.2. Generation of a sgRNA template by PCR.	103
3.2.3.1.3. Synthesis of sgRNA by in vitro transcription.	104
3.2.3.2. Genetic characterization of wild type zebrafish.	104
3.2.3.2.1. DNA extraction from adult zebrafish.	104
3.2.3.2.2. Amplification by PCR.	105
3.2.3.2.3. Sanger sequencing.	105
3.2.3.3. Microinjection of CRISPR-Cas9 components into zebrafish zygotes.	106
3.2.4. CRISPR-Cas9 editing efficiency analysis in individuals from the F0 generation.	106
3.2.4.1. DNA extraction from zebrafish embryos.	107
3.2.4.2. Amplification by PCR.	107
3.2.4.3. Heteroduplex analysis in acrylamide gels.	107
3.2.5. Genetic characterization of adult zebrafish from the F0 generation.	107
3.2.6. Generation of individuals from the F1 generation.	107
3.2.6.1. Genetic characterization of adult zebrafish from the F1 generation.	108
3.2.6.2. Allele identification in the F1 population.	108
3.2.6.3. Fragment analysis.	108
3.2.7. Generation of individuals from the F2 generation.	108
3.2.7.1. Characterization of individuals from the F2 generation.	108
3.2.7.2. Morphological analysis.	109
3.2.7.3. Behavioural analysis of the F2 generation using the light-dark test.	109
4. RESULTS	113
4.1. GENERATION AND CHARACTERIZATION OF AN <i>IN VITRO</i> MODEL TO STUDY THE FUNCTION OF <i>BTBD8</i>	113
4.1.1. Generation of a mutant hiPSC cell line for <i>BTBD8</i> using the CRISPR-Cas9 system.	113

4.1.1.1. CRISPR-Cas9 guide design and guide cloning into suitable vectors.....	113
4.1.1.2. Puromycin kill curve for the cell line Ctrl.BS-PBiPs37-Sv4F-1.....	116
4.1.1.3. Efficiency analysis of the knockout guides designed for <i>BTBD8</i>	117
4.1.2. <i>Characterization of the generated clones</i>	118
4.1.2.1. Allele identification.....	118
4.1.2.2. Karyotype analysis of the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-BTBD8-i2.....	120
4.1.2.3. Pluripotency marker expression in the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-BTBD8-i2.....	121
4.1.2.4. Trilineage differentiation of Ctrl.BS-PBiPs37-Sv4F-1-BTBD8-i2 edited cells.....	125
4.1.2.5. Off-target assessment of Ctrl.BS-PBiPs37-Sv4F-1-BTBD8-i2 edited cells.....	126
4.1.3. <i>Evaluation of the functional consequences of BTBD8 mutations in human brain organoids</i>	128
4.1.3.1. Assessment of alterations in brain organoid size.....	128
4.1.3.2. Immunohistochemistry characterization of human brain organoids harbouring mutations in BTBD8.....	130
4.1.4. <i>Evaluation of neurodevelopmental alterations associated with mutations in BTBD8 in cortical neurons derived from hiPSCs</i>	134
4.1.4.1. Characterization of cortical-like neurons derived from hiPSCs.....	134
4.1.4.2. Morphometric analysis of cortical neurons derived from <i>BTBD8</i> -mutant hiPSCs.....	135
4.1.4.3. Proteomic analysis of cortical neurons derived from <i>BTBD8</i> -mutant hiPSCs.....	137
4.2. GENERATION AND CHARACTERIZATION OF AN <i>IN VIVO</i> MODEL TO STUDY THE FUNCTION OF <i>FOSL2</i>	142
4.2.1. <i>Mortality after morpholine treatment</i>	142
4.2.2. <i>Knockdown verification in fosl2 morphants</i>	142
4.2.3. <i>Phenotypic characterization of fosl2 morphant embryos</i>	144
4.2.3.1. Morphometric analysis at 48 hfp.....	144
4.2.3.2. Cartilage staining of <i>fosl2</i> morphants.....	146
4.2.3.3. Behavioural response of the <i>fosl2</i> morphants using the light-dark test.....	147
4.3. GENERATION AND CHARACTERIZATION OF <i>IN VITRO</i> AND <i>IN VIVO</i> MODELS TO STUDY THE FUNCTION OF <i>TAOK1</i>	149
4.3.1. <i>Generation of a knockout model for TAOK1 in vitro</i>	149
4.3.1.1. CRISPR-Cas9 guide design, cloning into suitable vectors, and transfection.....	149
4.3.1.2. Guide efficiency analysis.....	152
4.3.2. <i>Characterization of the generated clones</i>	154
4.3.2.1. Allele identification.....	154
4.3.2.2. Karyotype analysis of the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2.....	156
4.3.2.3. Pluripotency maker expression in the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2.....	157
4.3.2.4. Trilineage differentiation of Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2 edited cells.....	161
4.3.2.5. Off-target assessment in the Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2 edited line.....	162
4.3.3. <i>Evaluation of the functional consequences of TAOK1 stop gain mutations in brain organoids</i>	165
4.3.3.1. Assessment of alterations in brain organoid size.....	165
4.3.3.2. Immunohistochemistry characterization of human brain organoids with knockout mutations in TAOK1.....	166
4.3.3.3. Transcriptomic analysis of <i>TAOK1</i> -mutant brain organoids by means of scRNA-seq.....	169
4.3.4. <i>Generation of an in vivo zebrafish model to study taok1</i>	176
4.3.4.1. Evaluation of the presence of <i>TAOK1</i> orthologous sequences in zebrafish.....	176
4.3.4.2. CRISPR-Cas9 guide design and synthesis.....	177
4.3.4.3. Characterization of the wild type individuals.....	179
4.3.4.4. CRISPR-Cas9 microinjection and guide efficiency analysis in the F0 generation of <i>taok1</i> -mutant zebrafish.....	180
4.3.4.5. Obtention of zebrafish from the F1 generation and genetic characterization.....	181
4.3.4.6. Obtention of zebrafish from the F2 generation and phenotypic characterization.....	182
4.3.4.6.1. Morphometric analysis of <i>taok1</i> mutant zebrafish.....	182
4.3.4.6.2. Behavioural response of the F2 larvae using the light-dark test.....	186
5. DISCUSSION	191
5.1. GENERATION AND CHARACTERIZATION OF <i>IN VITRO</i> MODELS TO STUDY THE FUNCTION OF <i>BTBD8</i>	191
5.1.1. <i>Functional implications of BTBD8's expression pattern and localization</i>	192
5.1.2. <i>Morphological alterations in Ctrl.BS-PBiPs37-Sv4F-1-BTBD8-i2 brain organoids</i>	192
5.1.3. <i>Interpretation of proteomic data from BTBD8-mutant neuronal cultures in the context of ASD</i>	193

5.2. GENERATION AND CHARACTERIZATION OF AN <i>IN VIVO</i> MODEL TO STUDY THE FUNCTION OF <i>FOSL2</i>	198
5.2.1. <i>Phenotypic alterations in fosl2 morphant zebrafish.</i>	199
5.3. GENERATION AND CHARACTERIZATION OF <i>IN VITRO</i> AND <i>IN VIVO</i> MODELS TO STUDY THE FUNCTION OF <i>TAOK1</i>	200
5.3.1. <i>Evaluation of the functional consequences of TAOK1 stop gain mutations in brain organoids.</i>	201
5.3.2. <i>Phenotypic alterations in taok1 mutant zebrafish.</i>	204
5.4. GENERAL CONSIDERATIONS REGARDING THE <i>IN VITRO</i> MODELS GENERATED IN THIS THESIS.	206
6. CONCLUSIONS	211
7. BIBLIOGRAPHY	215
SUPPLEMENTARY INDICES: FIGURES AND TABLES	235
FIGURE INDEX.....	237
TABLE INDEX.....	246
ANNEXES	249
ANNEXE I: APPROVAL OF THE ANIMAL EXPERIMENTATION ETHICS COMMITTEE.	251
ANNEXE II: AUTHORIZATION OF ANIMAL EXPERIMENTATION PROJECTS.	252
ANNEXE III: ANIMAL EXPERIMENTATION TRAINING CERTIFICATE.	254
ANNEXE IV: APPROVAL OF THE ETHICS COMMITTEE FOR THE PROJECT INCLUDING THE USE OF hiPSCs.	255
ANNEXE V: MATERIAL TRANSFER AGREEMENT (MTA) FOR THE USE OF THE hiPSC LINE CTRL. BS PBiPS37-Sv4F-1.	257
ANNEXE VI: SCIENTIFIC PUBLICATIONS INCLUDED IN THE THESIS.	260
ANNEXE VII: SUPPLEMENTARY PROTEOMIC DATA.	262

SUMMARY

SUMMARY

Autism Spectrum Disorders (ASD) are characterized by social and communicational impairments, in addition to restricted patterns of behaviour. Despite advances towards a better understanding of ASD's neurobiology, the mechanisms underlying these disorders are poorly understood.

For this reason, in this thesis *in vitro* and *in vivo* models were generated to study the implication of candidate ASD-risk genes in the development of the disorder. To do so, human induced pluripotent stem cells (hiPSCs) were selected as *in vitro* models. These are undifferentiated cells capable of self-renewal and differentiation into specialized cell types within an organism, including neurons and glia. Combining hiPSCs with the CRISPR-Cas9 system makes it possible to introduce specific genomic variants to replicate or correct point mutations associated with human diseases or disorders. Despite their advantages, cellular models have limited complexity and cannot replicate the core symptoms of ASD. For this reason, zebrafish (*Danio rerio*) was also proposed as an *in vivo* model to address the aforementioned limitations.

First, an *in vitro* model was generated and characterized to study *BTBD8*'s role in ASD. *BTBD8* encodes a protein with a BTB/POZ domain, potentially involved in transcriptional regulation and synaptic transmission, although the information regarding its role in these processes is limited. Here, CRISPR-Cas9 was employed to knockout the *BTBD8* in hiPSCs. As a result, a homogeneous cell population with biallelic mutations in this gene was isolated and characterized. Then, the edited cell line was differentiated into cortical neurons and brain organoids to study the impact of *BTBD8* dysfunction on neurodevelopment, which was analysed through immunostaining and proteomic approaches. Results obtained in this thesis support the functional involvement of *BTBD8* in the complex network of processes triggering alterations in neurodevelopment, including imbalances in neural progenitor cell proliferation and differentiation, as well as alterations in clathrin-mediated endocytosis, neurotransmitter exocytosis, vesicle transport, and synaptogenesis.

Second, *FOSL2* role in neurodevelopment was analysed. *FOSL2* belongs to a family of transcription factors involved in the regulation of cell proliferation, migration, survival, and differentiation in response to environmental signals. Recently, it was described a direct association between mutations in this gene and the presence of a syndromic developmental disorder with ASD. Here, a *fosl2* knockdown model was generated in zebrafish to study gene function during development. The obtained results indicate that *fosl2* expression is essential in early development stages in zebrafish.

Third, *in vitro* and *in vivo* models were generated for the study of the *TAOK1* function. *TAOK1* is a Ser/Thr kinase involved in the regulation of cellular signalling pathways. Here, an *in vitro* knockout model was generated in hiPSCs using the CRISPR-Cas9. The edited cell line was differentiated into human brain organoids, characterized through immunostaining and single-cell RNA sequencing (scRNA-seq). As a result, alterations associated with *TAOK1* dysfunction were observed in the proliferation of neural progenitor cells, neuronal migration, and neural differentiation. Overall, these data support the crucial regulatory role of *TAOK1* in the development of the human nervous system. A mutant zebrafish model was also generated to study the consequences *in vivo* of blocking the expression of *taok1*. Morphological and behavioural analyses of this model did not reveal significant differences between mutant individuals and controls. Nevertheless, it is necessary to extend the temporal scale of the study to determine if the lack of expression of the gene of interest causes morphological and locomotor alterations in later stages not covered in this thesis.

In summary, the generated models facilitate a deeper understanding of the molecular mechanisms guiding neurodevelopment and its alterations, thus leading to an improved diagnosis offered to patients and their families. In addition, these models open new avenues for the application of innovative therapeutical strategies capable of reversing or alleviating the patient's distinct symptoms. Consequently,

this approach signifies a paradigm shift towards personalized and precision diagnosis and therapy, a concept that was scarcely imaginable until just a few years ago.

RESUMO

Os trastornos do neurodesenvolvemento engloban un conxunto de afectacións cognitivas e do comportamento que comezan a manifestarse durante etapas temperás do desenvolvemento (dende etapas prenatais ata os primeiros anos de vida). En xeral, este tipo de trastornos van acompañados de alteracións a nivel neurolóxico que afectan as capacidades sociais, intelectuais, motoras e/ou lingüísticas dos individuos. Dentro deste conxunto de afectacións, atópanse os trastornos do espectro autista (TEA).

Os TEA engloban un conxunto de trastornos caracterizados pola presenza de alteracións a nivel social e comunicativo, que adoitan ir acompañados de intereses e condutas restritivas e/ou repetitivas. A nivel global, estímase que ao redor do 1 % da poboación mundial pode estar afectada por un trastorno deste tipo, aínda que esta porcentaxe podería ser moito máis elevada debido a limitacións diagnósticas asociadas a factores socioeconómicos.

Na actualidade, téñense identificado algúns factores xenéticos e ambientais asociados ao risco de padecer TEA, aínda que a meirande parte dos factores implicados no seu desenvolvemento continúan a ser incertos. Estimás recentes apuntan a que o 15 % dos casos de TEA poden estar relacionados cunha mutación xenética coñecida, pero espérase que esta porcentaxe aumente no futuro debido á aplicación diagnóstica de técnicas de secuenciación de nova xeración.

A nivel molecular os xenes asociados aos TEA están na súa meirande parte implicados na regulación da expresión xénica, comunicación neuronal, sinalización celular, modificacións post-transcripcionais e reorganización do citoesqueleto. En individuos con TEA, alteracións nestas funcións vincúlanse con anomalías na proliferación celular, apoptose, migración e sinaptoxénese. Ademais, as evidencias actuais apuntan a tres grandes mecanismos implicados na aparición dos TEA: perturbacións da migración neuronal, alteracións do equilibrio entre neuronas excitadoras e inhibitorias ou modificacións da morfoloxía neuronal. A nivel fisiolóxico, identificáronse varias rexións no sistema nervioso central (CNS) enriquecidas na expresión de xenes de risco de TEA, como o córtex, o cerebelo, a amígdala, o tálamo ou o corpo estriado.

A pesar dos numerosos avances cara a unha mellor comprensión da neurobioloxía do TEA, os mecanismos que subxacentes a estas alteracións son comprendidos en escasa medida. Por este motivo, o obxectivo central da presente tese de doutoramento consiste en avaliar se o emprego de modelos *in vitro* e *in vivo* é axeitado para o estudo deste tipo de trastornos.

O estudo do neurodesenvolvemento humano e as súas alteracións enfrontouse ao gran obstáculo de atopar un modelo axeitado para facelo. Durante moito tempo estes estudos fundamentáronse na análise de mostras humanas *post-mortem* e no uso de modelos animais, debido as complicacións técnicas, éticas e legais asociadas á obtención e emprego de cultivos primarios de cerebro humano. Este escenario cambiou coa optimización das tecnoloxías de reprogramación de células somáticas para xerar células nai pluripotentes inducidas (iPSCs). Esta nova tecnoloxía abriu a porta á xeración de modelos celulares proliferantes capaces de preservar a pegada xenética propia do individuo do que proceden.

O modelo *in vitro* seleccionado para o seu uso neste traballo son as células nai pluripotentes inducidas humanas (hiPSCs). As hiPSCs son células non especializadas que teñen capacidade de auto-renovación e diferenciación a outros tipos de células especializadas dun organismo,

incluídas células de liñaxe neuronal, as cales poden ser obtidas mediante cultivo en 2D ou formando estruturas 3D complexas, como os organoides cerebrais. Estas propiedades fan das hiPSCs un modelo moi valioso para o estudo do desenvolvemento *in vitro* dunha maneira accesible, o que resulta particularmente interesante para o estudo de condicións neurolóxicas. Ademais, a combinación do emprego de hiPSCs con tecnoloxías de edición do xenoma, tales como CRISPR-Cas9, fai posible introducir variantes xenómicas específicas para reproducir ou corrixir mutacións puntuais asociadas a enfermidades ou trastornos humanos.

Pese as súas vantaxes, os modelos celulares teñen unha complexidade limitada, non permitindo o estudo do impacto das mutacións a nivel dun organismo completo, nin tampouco o estudo do seu impacto no comportamento. Por este motivo, nesta tese de doutoramento tamén se propuxo empregar o peixe cebrá (*Danio rerio*) como modelo para abordar as limitacións previamente mencionadas. Este animal presenta notables vantaxes para o seu uso como modelo, tales como o seu pequeno tamaño e a robustez e manexabilidade da especie, que facilitan o seu mantemento nun laboratorio de maneira custe-eficiente. Ademais, o peixe cebrá acada a madurez sexual relativamente rápido (3 meses tras a fecundación), podendo chegar a ter postas semanais de 200 a 300 ovos. En relación a isto, esta especie ten fecundación externa e os embrións son transparentes, o que facilita en gran medida a manipulación e a análise de fenotipos de maneira non invasiva durante etapas temperáns do desenvolvemento. Ademais, ao redor do 75 % dos xenes involucrados en enfermidades humanas están conservados nesta especie, o que a fai un modelo moi útil para o estudo de enfermidades e trastornos de base xenética, tales como os TEA. O peixe cebrá tamén é capaz de manifestar comportamentos semellantes aos observados en individuos con este tipo de trastornos do neurodesenvolvemento, como a natación repetitiva, ou afectacións no grao de socialización entre conxéneres. A pesar das diferenzas existentes entre o desenvolvemento do sistema nervioso humano e o do peixe cebrá, ambas especies presentan unha organización semellante a nivel celular. O encéfalo de ambas especies presenta moitas rexións análogas, incluíndo o cerebelo, córtex, hipocampo e amígdala. Respecto aos tipos celulares neuronais, no peixe cebrá pódense atopar tanto neuronas como células de glia, incluídos astrocitos, oligodendrocitos, células de Schwann e microglía. Ademais, humanos e peixes cebrá comparten os principais neurotransmisores e circuitos neuronais asociados ao control de respostas básicas, como o sono e o sobresalto acústico, entre outros. Por todas estas razóns, o peixe cebrá é un modelo ideal para o estudo das causas e posibles tratamentos desta e doutras enfermidades neurolóxicas.

Tal e como se mencionou previamente, o obxectivo principal do presente traballo consiste en afondar nas bases xenéticas subxacentes aos TEA mediante a integración de modelos *in vitro* e *in vivo*. Nesta tese, seleccionáronse tres xenes de interese, *BTBD8*, *FOSL2* e *TAOK1*, en base a súa asociación previa con alteracións do neurodesenvolvemento e TEA, suxerida polos resultados de reportes clínicos. Sen embargo, en ningún dos tres casos se caracterizaron as alteracións a nivel molecular nin celular implicadas na aparición destes trastornos. Os obxectivos específicos da tese de doutoramento orientáronse á xeración de modelos axeitados para o estudo funcional de cada un dos xenes previamente citados.

En primeiro lugar, como parte desta tese de doutoramento xerouse e caracterizouse un modelo *in vitro* para o estudo da función do xene *BTBD8* (antes coñecido como *KIAA1107*). Debido a ausencia de ortólogos para este xene en peixe cebrá, este non puido ser estudado no modelo *in vivo* proposto.

BTBD8 codifica una proteína con dominio BTB/POZ, potencialmente implicada en funcións celulares críticas, tales como a regulación transcricional e a transmisión sináptica, aínda que é pouca a información referente ao seu papel nestes procesos. Neste traballo, empregouse o sistema de edición xenética CRISPR-Cas9 para bloquear a expresión do xene

BTBD8 en hiPSCs. Unha vez illada unha poboación celular homoxénea con mutacións bialélicas neste xene (que foi denominada Ctrl.BS-PBiPS37-SvF4-1-*BTBD8*-i2), leváronse a cabo ensaios de caracterización para corroborar a ausencia de alteracións a nivel xenómico como consecuencia do proceso de edición. En concreto, avalíouse a ausencia de alteracións a nivel do cariotipo, así como a ausencia de eventos de edición inespecíficos no xenoma das células. Ademais, comprobouse que a liña celular obtida mantíña a pluripotencia, é dicir, a capacidade de diferenciación a tipos celulares dos tres liñaxes embrionarios.

Unha vez caracterizada, a liña de hiPSCs Ctrl.BS-PBiPS37-SvF4-1-*BTBD8*-i2 foi diferenciada a neuronas corticais e organoides de cerebro para estudar o impacto das mutacións incorporadas en *BTBD8* no neurodesenvolvemento. Por unha banda, os resultados obtidos no presente estudo indican que *BTBD8* se localiza principalmente no núcleo celular, ao contrario do que acontece en cultivos primarios de neuronas procedentes de modelos murinos, onde a localización é maioritariamente citoplasmática. Estas observacións poderían estar relacionadas coa inmadurez dos cultivos neuronais xerados a partir de hiPSCs, pero tamén poderían indicar que a localización subcelular de *BTBD8* difire entre especies. Esta última hipótese suxeriría un importante papel evolutivo de *BTBD8* na formación do sistema nervioso, aínda que é preciso realizar estudos adicionais que permitan confirmar estas hipóteses. Por outra banda, o seguimento do desenvolvemento de organoides cerebrais humanos derivados da liña Ctrl.BS-PBiPS37-SvF4-1-*BTBD8*-i2 revelou que estes presentan un tamaño significativamente maior que os seus controis isoxénicos. Este fenotipo de macrocefalia, xunto coas alteracións nas rutas de apoptose e de degradación de proteínas observados no estudo proteómico realizado, suxiren un desequilibrio entre a proliferación e diferenciación dos proxenitores neurais asociado á disfunción de *BTBD8*. Ademais, os cultivos neuronais derivados da liña Ctrl.BS-PBiPS37-SvF4-1-*BTBD8*-i2 mostran sinais de alteracións na diferenciación neuronal, incluíndo unha expresión elevada de marcadores de células proxenitoras, unha expresión reducida de marcadores neuronais GABAérxicos e unha diferenciación retardada de neuronas corticais excitadoras. Todo isto que podería levar a un retraso no desenvolvemento, así como a un desequilibrio no número de neuronas excitadoras e inhibitorias, un dos mecanismos que se postulan como principais implicados na patoxénese dos TEA. Ademais, as neuronas corticais derivadas da liña Ctrl.BS-PBiPS37-SvF4-1-*BTBD8*-i2 mostran tamén alteracións na ruta de endocitose mediada por clatrina, na exocitose de neurotransmisores, no transporte de vesículas e na sinaptoxénese, o que suxire alteracións na excitabilidade das células como consecuencia da disfunción de *BTBD8*. Estas alteracións no sistema endosomal poden levar a unha cascada de eventos que inflúen no desenvolvemento dos TEA, afectando a numerosos mecanismos moleculares, incluídos a regulación da expresión xénica, a transmisión sináptica, a sinalización celular e a remodelación do citoesqueleto. Con todo, a complexa interacción entre os xenes regulados polo sistema endosomal e aqueles que gobernan a súa función non foi totalmente descuberta polo momento.

O modelo xerado para o estudo de *BTBD8* é, polo que sabemos, o primeiro modelo *in vitro* humano xerado para estudar a función deste xene. Ademais, os resultados presentados apoian a implicación funcional de *BTBD8* na complexa rede de procesos desencadeantes de alteracións no neurodesenvolvemento, compatibles con alteracións previamente asociadas aos TEA.

En segundo lugar, neste traballo propúxose estudar a función do xene *fosl2* no neurodesenvolvemento mediante a xeración dun modelo de silenciamento mediante morfólino en peixe cebra. O xene humano *FOSL2* pertence a unha familia de factores de transcrición implicados na regulación da proliferación celular, migración, supervivencia e diferenciación en resposta a sinais ambientais. En decembro de 2022, Cospain e colaboradores describiron por primeira vez unha asociación directa entre mutacións en *FOSL2* e a presenza dun trastorno do

desenvolvemento síndrome con TEA. Estes autores identificaron un aumento da estabilidade dos complexos de transcrición AP-1, formados por *FOSL2* e *JUN*, como potencial mecanismo desencadeante do fenotipo observado. Non obstante, ata o momento, o papel de *FOSL2* na función do sistema nervioso non foi explorado máis afondo. Ademais, Kubra e colaboradores exploraron o potencial do peixe cebrá para estudar os xenes da familia *fos*. Nos mamíferos, identificáronse catro membros desta familia (*C-FOS*, *FOSB*, *FOSL1* e *FOSL2*), mentres que no peixe cebrá están presentes seis, xa que dous deles foron duplicados durante a evolución (*fosaa/fosab*, *foslla/fosl1b*, *fosb* e *fosl2*). De maneira notable, *FOSL2* ten un só xene ortólogo no peixe cebrá, coñecido como *fosl2*, o que simplifica o estudo da súa función. Ademais, *fosl2* é o membro desta familia que presenta uns maiores niveis de expresión cerebral, o que podería indicar un papel importante na regulación do desenvolvemento do sistema nervioso central, así como unha función conservada entre vertebrados.

O modelo de silenciamento de *fosl2* en peixe cebrá proposto na presente tese de doutoramento pretende abordar a utilidade deste modelo animal para o estudo da función de *fosl2*, así como avaliar o impacto fenotípico que ten a redución da expresión deste xene. Os resultados obtidos neste traballo indican que, as 48 horas tras a fecundación, os embrións de peixe cebrá coa expresión de *fosl2* silenciada presentan unha diminución do tamaño corporal, cefálico e ocular en comparación cos individuos non tratados. Ademais, a inactivación de *fosl2* parece provocar un incremento do ratio entre a cabeza e o corpo dos individuos, así como unha redución do ratio entre o ollo e a cabeza. Tendo en conta que en estadios temperáns do desenvolvemento de peixe cebrá, o tamaño da cabeza é relativamente maior que o tamaño do corpo, estes resultados apuntan a un retraso no desenvolvemento asociado coa redución da expresión de *fosl2*, apoiando o seu rol regulador neste proceso. Ademais, a elevada taxa de mortalidade dos embrións tratados con morfólino para silenciar *fosl2* suxire un posible fenotipo letal asociado á redución da expresión deste xene, o que indica unha función crítica durante o desenvolvemento temperán.

Tendo en conta que *fosl2* parece estar directamente implicado na formación dos tecidos óseos e cartilaxinosos, no modelo xerado avalíouse a presenza de alteracións na formación da mandíbula as 120 horas tras a fecundación. Os resultados desta análise non revelaron alteracións significativas na formación desa estrutura, aínda que isto podería deberse a unha redución do efecto do morfólino empregado tras 5 días. A inactivación de *fosl2* tampouco parece influír na locomoción dos individuos as 120 horas tras a fecundación, o cal podería deberse tamén ao motivo previamente citado. Ensaio futuros realizados con liñas mutantes estables para *fosl2* poderían ser de utilidade para confirmar os fenotipos observados ás 48 horas tras a fecundación, así como para confirmar se a disfunción de *fosl2* está ou non implicada na presenza de alteracións na formación do esqueleto e na locomoción dos individuos.

O terceiro e último obxectivo desta tese de doutoramento foi a xeración e caracterización de modelos *in vitro* e *in vivo* para o estudo da función do xene *TAOK1* e o seu papel no desenvolvemento do sistema nervioso. *TAOK1* é un dos membros da subfamilia de proteínas quinases de mil un aminoácidos (TAOKs) e está implicado na regulación de vías de sinalización celulares. Na actualidade, existen múltiples reportes da presenza de mutacións neste xene en individuos diagnosticados con algún tipo de trastorno do neurodesenvolvemento, incluídos TEA con diversos graos de severidade. A pesar de ser bastante variable, varios estudos identificaron síntomas comúns presentes en pacientes con mutacións en *TAOK1*, como o retraso no desenvolvemento, a discapacidade intelectual, macrocefalia, dificultades para a alimentación, hipotonía, problemas de comportamento, e laxitude articular. Curiosamente, as variantes xenómicas en *TAOK1* asociadas a trastornos do neurodesenvolvemento foron vinculadas tanto con incrementos como con reducións dos niveis de expresión deste xene,

causando alteracións na morfoloxía e migración neural. Isto suxire un papel esencial de *TAOK1* na regulación rigorosa da formación do sistema nervioso. Pese a isto, poucos son os estudos nos que profundara sobre a implicación de *TAOK1* nestes procesos.

Co fin de abordar esa falta de coñecemento, neste traballo xerouse un modelo *in vitro* en hiPSCs para estudar os efectos do bloqueo da expresión deste xene nun modelo celular humano, sendo esta a primeira vez que este abordaxe se aplica para o estudo de *TAOK1*. O sistema de edición xenética CRISPR-Cas9 foi empregado para bloquear a expresión de do xene de interese en hiPSC controis. A continuación, illouse unha poboación celular homoxénea con mutacións bialélicas truncantes neste xene (que foi denominada Ctrl.BS-PBiPS37-SvF4-1-*TAOK1*-i2) e leváronse a cabo ensaios de caracterización tal e como foi previamente descrito para o estudo de *BTBD8*.

A caracterización inmunohistoquímica de organoides cerebrais obtidos a partir da liña editada Ctrl.BS-PBiPS37-SvF4-1-*TAOK1*-i2, combinada cos datos de secuenciación de ARN a partir de célula única (scRNA-seq), permitiu a identificación de variacións fenotípicas asociadas a falta de expresión de *TAOK1*. Por unha banda, a macrocefalia observada nos organoides cerebrais Ctrl.BS-PBiPS37-SvF4-1-*TAOK1*-i2, xunto coas alteracións nas poboacións de células proxenitoras neurais, suxiren un desequilibrio entre a proliferación e a diferenciación deste tipo de células asociado ao bloqueo da expresión de *TAOK1*. Ademais, os organoides cerebrais mutantes para este xene presentan defectos na migración neuronal e na formación das láminas corticais, tal e como suxiren a redución no número de células Cajal-Retzius e das poboacións que expresan xenes de síntese da matriz extracelular. Os organoides cerebrais Ctrl.BS-PBiPS37-SvF4-1-*TAOK1*-i2 mostran unha diferenciación neural aberrante que afecta tamén a astrocitos, neuronas excitadoras e inhibitoras. En conxunto, estes datos apoian a importante función reguladora de *TAOK1* durante o desenvolvemento do sistema nervioso humano. Ademais, as alteracións ligadas a falta de expresión deste xene son comúns en casos de TEA, tal e como se mencionou anteriormente, o que apoia a idoneidade do modelo xerado para o estudo destes trastornos e do papel de *TAOK1* no seu desencadeamento.

Ademais de xerar un modelo *in vitro* a partir de células humanas para o estudo de *TAOK1*, como parte desta tese doutoral xerouse un modelo mutante de peixe cebra para estudar as consecuencias *in vivo* do bloqueo da expresión do ortólogo deste xene, denominado *taok1*. As análises morfolóxicas realizadas sobre este modelo non revelaron diferenzas significativas entre individuos mutantes e controis. Resultados obtidos nesta tese tamén apuntan a que a presenza de mutacións truncantes en heterocigose poden estar ligadas a unha menor actividade locomotora no peixe cebra as 96 horas tras a fecundación. Con todo, o desenvolvemento do modelo de peixe cebra só foi analizado durante as primeiras 96 horas tras a fecundación. Por este motivo, é preciso ampliar a escala temporal do estudo para poder determinar se a falta de expresión do xene de interese provoca alteracións morfolóxicas e locomotoras en estadios máis avanzados, así como ampliar os estudos para determinar se as alteracións celulares observadas en organoides cerebrais se replican neste modelo animal.

A xeración de modelos *in vitro* e *in vivo* que repliquen variantes xenéticas asociadas aos trastornos do neurodesenvolvemento, incluíndo os TEA, promete proporcionar unha comprensión máis profunda dos mecanismos moleculares subxacentes a estas alteracións. O coñecemento xerado como resultado destes estudos non só mellorará a nosa comprensión do complexo proceso de neurodesenvolvemento humano, senón que tamén poderá ser aplicado á mellorara da calidade de vida das persoas con TEA, pois revertera nun mellor diagnóstico e consello xenético. Ademais, estes modelos poderán ser empregados para o desenvolvemento e optimización novas opcións terapéuticas. Na actualidade, hai varios fármacos dispoñibles para tratar síntomas secundarios asociados a este trastorno, pero ningún deles esta destinado

especificamente a tratar os síntomas centrais dos TEA, polo que o seu impacto na calidade de vida dos afectados é limitado. Os modelos xerados na presente tese de doutoramento, tanto *in vitro* como *in vivo*, poderían ser aplicables nun futuro próximo ao desenvolvemento de novas opcións terapéuticas avanzadas no camiño cara unha medicina personalizada de precisión.

INTRODUCTION

1. INTRODUCTION

1.1. NEURODEVELOPMENTAL DISORDERS.

According to the 11th revision of the International Classification of Diseases by the World Health Organization (ICD-11, WHO) and the 5th Diagnostic and Statistical Manual of Mental Disorders from the American Psychiatric Association (DSM-5, APA), neurodevelopmental disorders are a group of behavioural and cognitive conditions with onset on the developmental period. They are characterized by alterations in brain processes that involve significant difficulties in social, intellectual, motor or language performance. However, the severity of these symptoms is variable, and sometimes it is complicated to define the boundaries with typical development^{1,2}. Whenever possible, clinicians use standardized tests that take into account cultural and age biases to diagnose this type of disorders¹.

Oftentimes, the aetiology of neurodevelopmental disorders is unclear. In some cases, they can be associated to injuries in the central nervous system (CNS) during development, but environmental and genetic factors could also be linked to their pathogenesis¹.

According to the WHO's ICD-11, neurodevelopmental disorders include: autism spectrum disorders, disorders of intellectual development, developmental speech or language disorders, developmental learning disorders, developmental motor coordination disorders and attention deficit hyperactivity disorders, among others¹.

1.2. AUTISM SPECTRUM DISORDERS.

Autism Spectrum Disorders (ASD) are characterized by the persistent presence of social and communicational impairments, in addition to restricted, repetitive and/or inflexible patterns of behaviour, interests, or activities. The symptoms must be atypical for the age and sociocultural context of the individual, but their severity can vary between cases, hence the term spectrum¹⁻³. ASD onset occurs during development, yet the symptoms change over time and might even be masked in some cases^{1,2}.

The current consensus establishes that ASD include disorders previously diagnosed as early infantile autism, childhood autism, Kanner's autism, high-functioning autism, atypical autism, pervasive developmental disorder not otherwise specified, childhood disintegrative disorder, and Asperger's disorder².

1.2.1. Prevalence of ASD.

The global prevalence of ASD is 1 %, nevertheless it is important to mention that it varies worldwide, ranging from 0.01 % to 4.36 %, according to recent data³⁻⁵. Also, it is important to consider that some cases might not have been taken into account in the reported prevalence due to underdiagnosis, which increases depending on demographic and socioeconomical factors^{2,3,6}. Figure 1 shows the worldwide ASD prevalence, based on the most recent available data up to date (2019)^{7,8}.

Prevalence studies are consistent in highlighting that ASD are more frequent in males than in females, with current data indicating that the male-to-female ratio is 4.2. However, it is also

important to notice that females diagnosed with ASD usually manifest symptoms that are more severe than males. This evidence has led to the “female protective effect” theory, that claims that more factors are needed in females to manifest an impairment equivalent to males. However, the underlying mechanisms are still to be discovered^{3,4,6,9}.

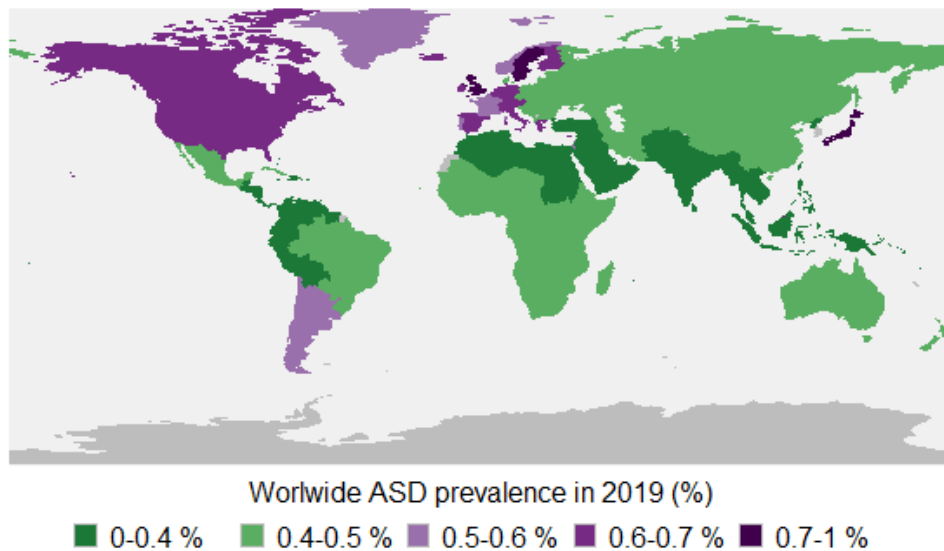


Figure 1. Worldwide ASD prevalence in 2019. Dark green: 0-0.4 %; light green: 0.4-0.5 %; lavender: 0.5-0.6 %; violet: 0.6-0.7 %; dark violet: 0.7-1 %; grey: no data available. The figure was created by the author using R software (R Core Team, Vienna, Austria) to represent open access data, previously standardized by age and sex^{7,8,10}.

1.2.2. Aetiology of ASD.

ASD cases can be broadly classified into two main categories: syndromic and non-syndromic. Syndromic forms of ASD have a well-characterized aetiology, whereas the origins of the non-syndromic forms remain unknown. Despite most ASD cases being categorised as non-syndromic (> 90 %), it is known that both environmental and genetic factors play an important role in ASD development, as pointed out by twin studies. Twin studies are based on the comparison of the phenotypes of monozygotic and dizygotic twins to study the heritability of a certain character. Since monozygotic twins share the 100 % of their genomes, phenotypic differences between them must be due to environmental factors. However, in dizygotic twins the same character is going to be influenced by both genetic and environmental factors. For ASD, the results obtained from twin studies are quite variable, but the concordance between monozygotic twins is consistently higher than between dizygotic twins, which indicates the high heritability of the disorder (> 50 %)^{5,6,11–16}.

ASD are complex disorders which onset is, in general, the consequence of the combination of several risk factors. As an example, some environmental factors can induce epigenetic changes and/or modify *de novo* mutation rates, which directly impacts gene expression during neurodevelopment^{5,6,17}.

1.2.2.1. Environmental risk factors.

Despite differences in estimates between studies, all conclude that ASD variance is influenced by environmental factors. It has been shown that advanced parental age and maternal

health during pregnancy are associated to ASD, as well as birth-related complications such as ischemia, hypoxia, among others. Early exposure to heavy metals -like mercury and lead-, and drugs -such as valproate and thalidomide- has also been addressed as a potential ASD risk factor^{5,6,12–15,18}. However, it is extremely challenging to determine whether a certain environmental factor is directly related to ASD development or whether the association is non-causative, especially if the studies are based on observations, which is the case of many of the current publications on this field^{5,6,11,12,16,17}.

1.2.2.2. Genetic risk factors.

The term genetic variation refers to alterations in the DNA sequence or chromosomal structure among individuals. Genetic variation can be classified according to the variant's heritability (inherited/*de novo*) and frequency (common/rare). Generally, rare genetic variation has a low frequency in the population due to selection (minor allele frequency, MAF < 1 %), but their impact on the onset of disease is greater than that of common variants. *De novo* variants occur for the first time in an individual and, in general, have a high impact on the onset of disease^{5,19}.

The identification of all the genetic variants that contribute to ASD development has been the focus of multiple studies across the years, which will be summarized in this section.

1.2.2.2.1. Technologies applied for the identification of ASD's genetic risk factors.

The first attempts to identify ASD-associated genetic variants were conducted with the help of linkage studies. These studies are based on the analysis of the segregation pattern of genetic markers (microsatellites or SNPs) across the genome. Despite being helpful to identify chromosomal regions associated with the disorder, most of the obtained results were not reproducible due to the heterogeneity of the selected cohorts of patients^{20,21}.

The next approach that was used to study ASD-genetic variation was based on the study of selected variants (SNPs) in candidate genes, based on their function or localization in previously identified ASD-associated chromosomal regions. Although very promising in the beginning, the results of these studies were also very hard to replicate²⁰.

The next big step for the study of ASD-associated genetic variation came along with the development of genome wide association studies (GWAS). Briefly, GWAS consists in the study of common variants (SNPs) across the whole genome to identify markers that differ from case and control groups. In comparison with the previous approaches, GWAS rely on bigger cohorts of individuals which improves the quality and reproducibility of their results. Additionally, novel bioinformatic pipelines were used for GWAS analysis, which allowed researchers to identify multiple SNPs and genes significantly associated with ASD in a more reproducible way^{19,20,22}.

Although promising, GWAS are limited by the number of variants selected to be analysed. This limitation was overcome thanks to the development of next-generation sequencing techniques (NGS), which are the base of whole-exome sequencing (WES) and whole-genome sequencing approaches (WGS). Both techniques allow the identification of *de novo* loss-of-function mutations, that have the potential to perturb the activity of the gene in which they are located and for that reason are classified as potentially pathogenic. Loss-of-function mutations can appear in the reproductive cells (germinal mutations) or in the first divisions of the zygote (postzygotic mutations, PZMs)^{19,22,23}. Until recently, studies of ASD-genetic variation were

only focused on the role of *de novo* germinal mutations, disregarding the implication of PZMs on its pathogenesis, due to the lack of novel bioinformatic tools to identify them. To date, it is estimated that 3-4 % of ASD's genetic risk is associated with *de novo* PZMs that affects genes expressed in the CNS²⁴⁻²⁶. WES and WGS also made possible to identify structural variants associated with ASD, including small insertions, deletions, and copy number variations (CNVs), which were hard to detect with other technologies^{19,22,23}.

Most part of NGS-based analysis applied to the study of ASD have been conducted by WES, which is limited to the sequencing of the coding regions of the genome, since it was developed earlier than WGS²⁷⁻²⁹. The main difference between WES and WGS is that the first one is limited to the coding regions of the genome, while WGS can be used to detect all types of genetic variation in coding and non-coding regions. For this reason, it is expected that WGS-based analyses contribute to a better understanding of ASD pathogenesis and its underlying mechanisms in the near future^{22,23}.

1.2.2.2.2. Databases of ASD genetic risk factors.

According to recent data, around 15 % of ASD cases can be linked to a known genetic mutation, but this percentage is expected to increase in the future in part due to the application of NGS to its diagnosis^{2,5,6,19}. However, it is also important to considered that genetic mutations linked to ASD are not fully penetrant and, in most cases, ASD risk is determined by many *loci*^{2,5,19}.

There are several databases that recollect information about the mutations that have been reported as associated with ASD in the literature, such as Autism Genetic Database (AGD)³⁰, AutismKB 2.0³¹, SFARI Gene³² and AutDB³³. A substantial period has passed without any updates to the AGD and AutismKB 2.0 databases, but both AutDB and SFARI gene keep updating the available information regularly (last updates from January and April 2023, respectively). Since the SFARI gene database is the one with the most recent updated, it was chosen to be explored more into detail on this introduction.

In the SFARI gene database, genes are classified according to a score that indicates the level of support of the association between mutations in the gene and ASD development (see Table 1 for further details on the SFARI gene score classification criteria).

Table 1. SFARI gene database scoring system criteria. Information obtained from SFARI gene database (April 2023 release)³².

SFARI GENE SCORE	CRITERIA
Score 1 ("high confidence")	Reported <i>de novo</i> gene likely disrupting mutations: >3. Genome-wide significant, false discovery rate <0.1.
Score 2 ("strong candidate")	Reported <i>de novo</i> gene likely disrupting mutations: 2. Significant in one GWAS/consistently replicated with a functional effect.
Score 3 ("suggestive evidence")	Reported <i>de novo</i> gene likely disrupting mutations: 1. Significant, but non-replicated studies.
Syndromic	Mutations with increased ASD-risk, but not exclusively.

In the latest SFARI gene database release, from April 2023, a total of 17 CNVs were included as associated with ASD, with almost all of them (n = 14) having more than 100 reports of association with the disorder. The SFARI gene database also included information about common and rare genetic variants linked to ASD, with a total of 1125 genes included in the last release (1123 protein coding, 2 lncRNAs, 1 miRNA and 2 processed pseudogenes), which represents more than a 5 % of the total number of nuclear protein coding genes in humans. Most

of them are linked to non-syndromic forms of ASD and classified as strong candidates (Figure 2).

As mentioned earlier, there is a great number of genes identified as potentially linked to ASD development in public databases. However, most of these findings lack functional validation. Conducting functional studies to investigate the role of these genes in ASD development can not only enhance our understanding of the molecular basis underlying ASD, but also untangle the complexity of human neurodevelopment.

In order to contribute to this process, this thesis focuses on the study of three genes potentially linked to ASD, which will be addressed in the following subheadings.

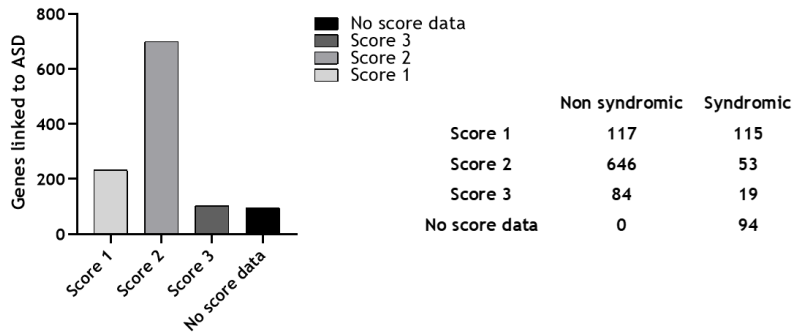


Figure 2. Number of genes linked to ASD according to the SFARI gene database (April 2023 release)³², distributed by their ASD-association score.

In terms of genomic distribution, ASD-associated genes have been identified in all human chromosomes, but it is remarkable their abundance in chromosome X (Figure 3), which could be linked to the male-to-female ASD ratio previously mentioned.

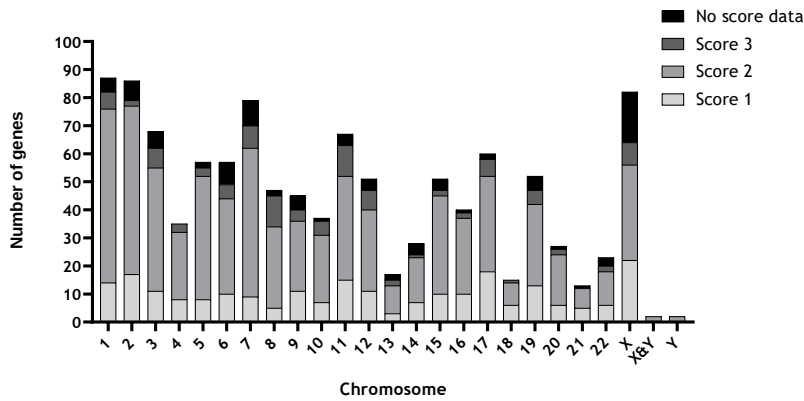


Figure 3. Protein coding genes associated with ASD according to the SFARI gene database (April 2023 release)³². Genes are distributed according to the chromosome where they are located. On the left, the total number of genes is indicated. On the right, the number of genes linked to ASD was corrected by the total number of protein coding genes on each chromosome.

1.2.3. Molecular pathways altered in ASD.

As exposed in the previous section, ASD are characterized by a great genetic variability, with hundreds of genes potentially involved on its development. In this complex scenario, several groups have been focused on the application of systems biology for the identification of underlying mechanisms that could explain how this huge genetic variation converges in ASD's phenotype.

As a result, several regions in the CNS have been found enriched in the expression of ASD-risk genes. One of those regions is the cortex, which was expected considering its role in cognition. Other CNS regions enriched in the expression of ASD-associated genes are the cerebellum, amygdala, thalamus, and striatum. In addition, the expression of most ASD-associated genes has been detected during foetal development, in both immature and mature neurons^{16,27,34–37}.

At the molecular level, in general ASD-associated genes fall into two main functional categories: regulation of gene expression and neuronal communication. Although not as frequent, genes involved in cell signalling, post-translational modifications and cytoskeleton rearrangements have also been linked to ASD development. In this sense, despite technical limitations, alterations in these functions have been linked to common abnormalities in individuals with ASD, including alterations in cell proliferation, apoptosis, migration, and synaptogenesis. Due to alterations in these functions, ASD pathogenesis apparently occurs. To date, evidence points out to three main mechanisms that could underlie ASD phenotype. The first one is the alteration of cell migration that could be mediated by the reelin pathway. The second one is the imbalance between excitatory (glutamatergic) and inhibitory (GABAergic) neurons. The third one is the alteration in synapse formation and dendritic morphology^{16,27,35,36,38–42}.

Despite advances towards a better understanding of the neurobiology of ASD, the precise mechanisms underlying these alterations are poorly understood. Improving our knowledge about these mechanisms not only could be helpful towards the development of more specific therapeutical options to treat core ASD symptoms, but also to better understand the functioning of the complex human CNS.

1.2.4. Diagnosis of ASD.

As previously mentioned, ASD onset happens during development when symptoms start to be apparent, and the disorder can be diagnosed. In general, this occurs around the 18-24 month of life, but it depends on the severity of the symptoms^{2,3}.

ASD cases are diagnosed based on clinical criteria, as indicated by the WHO's ICD-11 and the APA's DSM-5. Both manuals agree on the four requirements that are needed for an ASD diagnosis to be established, which are: persistent deficits in social communication and interaction; persistent restrictive, repetitive, and inflexible pattern of behaviour, interests, or activities; onset during the developmental period; and significant impairments in multiple areas of functioning caused by the symptoms. Also, the observed symptoms cannot be better explained by another disorder and the manifestations can vary during the lifespan of an individual^{1,2}.

The DSM-5 also considers the need to add several specifiers to the ASD diagnosis, to better describe each single case. These specifiers are related to the severity of the symptoms, which are related to the level of support needed by the individual; the presence of intellectual and/or language impairments; the association to a known risk factor, that can be genetic, medical, or environmental; and the presence of catatonic-like motor behaviour².

Around 70 % of ASD cases have one comorbid mental disorder or medical condition, while the 40 % have two or more. The most frequent ones are language problems, intellectual disability, and ADHD, but there is a great variability among cases^{2,6,43,44}.

1.2.5. Treatment of ASD.

Table 2. Pharmacological treatments that have been tested to alleviate ASD symptoms.

TYPE OF DRUG	NAME	AREAS WITH IMPROVEMENT	POTENTIAL SIDE EFFECTS
Atypical antipsychotic	Risperidone	Irritability, socialization, behaviour, hyperactivity, and communication. Specifically approved by the FDA for ASD-treatment.	Weight gain and somnolence.
	Aripiprazole	Irritability. Specifically approved by the FDA for ASD-treatment.	Weight gain, somnolence, tremor, and drooling.
	Olanzapine	Irritability, anxiety, socialization, and communication.	Weight gain and loss of strength.
	Ziprasidone	Irritability, agitation, and aggression.	Somnolence.
	Lurasidone	Small global improvement.	Vomiting and somnolence.
	Paliperidone	Irritability.	Weight gain and extrapyramidal symptoms.
Typical antipsychotic	Haloperidol	Hyperactivity, behaviour, and learning.	Somnolence, irritability, and acute dystonic reactions.
Selective Serotonin Reuptake Inhibitors (SSRIs)	Sertraline	Irritability, anxiety, aggression, and behaviour.	Not significant.
	Escitalopram	Irritability and global improvement.	Hyperactivity and irritability.
Serotonin-norepinephrine reuptake inhibitor (SNRI)	Venlafaxine	Behaviour, socialization, and communication.	Inattention, nausea, and polyuria.
	Atomoxetine	Hyperactivity and impulsivity.	Not significant.
Noradrenergic and specific serotonergic antidepressant (NaSSA)	Mirtazapine	Global improvement.	Increased appetite, irritability, and somnolence.
Tricyclic antidepressant	Clomipramine	Hyperactivity and behaviour.	Worsening symptoms of ASD.
	Desipramine	Hyperactivity.	Irritability and aggression.
Mood stabilizer	Divalproex sodium	Behaviour.	Irritability.
Psychostimulant	Methylphenidate	Hyperactivity.	Decreased appetite, insomnia, irritability, social withdrawal, and emotional outbursts.
Alpha-2 adrenergic receptor agonist	Clonidine	Hyperactivity and socialization.	Somnolence.
	Guanfacine	Hyperactivity and impulsiveness.	Somnolence and decreased appetite.
Opioid antagonist	Naltrexone	Hyperactivity and impulsiveness.	Not significant.
Reversible acetylcholinesterase inhibitor	Galantamine	Irritability, hyperactivity, and socialization.	Not significant.
Dopaminergic agent	Amantadine	Hyperactivity, irritability, and socialization.	Nausea, dizziness, and insomnia
Gamma-aminobutyric acid (GABA) agonist	Arbaclofen	Global improvement.	Emotional outbursts and sedation.
Hormone	Oxytocin	Socialization and behaviour.	Headache, gastrointestinal alterations, anxiety.
	Melatonin	Sleep disturbances.	Not significant.
Vasopressin V1a receptor antagonist	Balovaptan	Socialization and communication.	Not significant.

As it was previously mentioned, in the most severe cases of ASD, the symptoms can be incredibly disruptive for the daily functioning of the individual. Nowadays, there are several therapies available to ameliorate the symptoms of ASD, with the objective of reducing their impact on the individual's quality of life. However, none of them can fully alleviate the symptoms of ASD, neither alone nor combined, and their effect is limited and variable among different cases^{11,45–47}.

Several therapies, both pharmacological and non-pharmacological, have been tested to treat the symptoms of ASD, which are synthesized on Table 2 and Table 3, respectively. The drugs indicated on Table 2, except for arbaclofen and balovaptan, are approved by both the United States Food and Drug Administration (FDA) and the European Medical Administration (EMA) for the treatment of diverse diseases and disorders. However, up to date, only two of them (risperidone and aripiprazole) have been indicated by the FDA for the treatment of irritability associated with ASD. The EMA, however, does not indicate any of them for ASD-treatment since they do not target the core symptoms of the disorder. Regarding balovaptan, it received the Breakthrough Therapy Designation by the FDA for ASD-treatment due to initial promising results. However, different clinical trials conducted with the same compound showed very different outcomes, with no proven efficacy for ASD-treatment in children^{11,45–48}.

Table 3. Non-pharmacological therapies that have been used to alleviate ASD symptoms.

THERAPY	PROCEDURE	AREAS WITH IMPROVEMENT
Cognitive-behavioural therapy (CBT)	Psychoeducational techniques based on worksheets and visual aids.	Socialization and emotional regulation.
Social-behaviour therapy (SBT)	Applied behaviour analysis (ABA), Picture Exchange Communication System (PECS), etc.	Independence, communication, and socialization.
Music therapy	Use of music to impact the neural connectivity in the developing brain.	Socialization and communication.
Complementary and integrative health	Use of natural products and mind/body interventions.	Dependent on the individual.
Diet modifications	Oral supplementation.	Gastrointestinal alterations, which are common in ASD.

1.3. GENOME EDITING TOOLS IN THE STUDY OF ASD'S GENETIC VARIATION.

For a long time, studying the functional contribution of a gene in human neurodevelopment was out of reach for the scientific community. However, this scenario began to change with the development of genome editing tools that allowed scientists to specifically modify single genes to study their role in health and disease.

To date, there are four main types of genome-editing tools, which will be discussed in the following sections: meganucleases, zinc finger nucleases (ZFNs), transcription activator like effector nucleases (TALENs) and CRISPR-Cas9 systems. All these systems were developed from the same idea: the creation of double-strand breaks (DSBs) into a DNA sequence activates DNA repair mechanisms in eukaryotic cells and researchers learned to take advantage of these intrinsic mechanisms to alter DNA to their convenience⁴⁹.

There are two main DNA repair pathways in eukaryotic cells: non-homologous end joining (NHEJ) and homology-directed repair (HDR), represented in Figure 4. The first acts as a fast mechanism to eliminate DSBs, but it tends to generate insertions and deletions (indels) of nucleotides because of its activity. Therefore, reading frames might be altered or early stop codons inserted in a DNA sequence, compromising the function of the coded protein. However, for researchers this can be a useful tool to generate knockouts (KO) in genes of interest. The second repair mechanism, HDR, is a fine-tuned correction system that corrects DSBs by using

another DNA strand as a template. This mechanism can be tricked when an external template with the desired DNA modification is added to the cells^{49–51}.

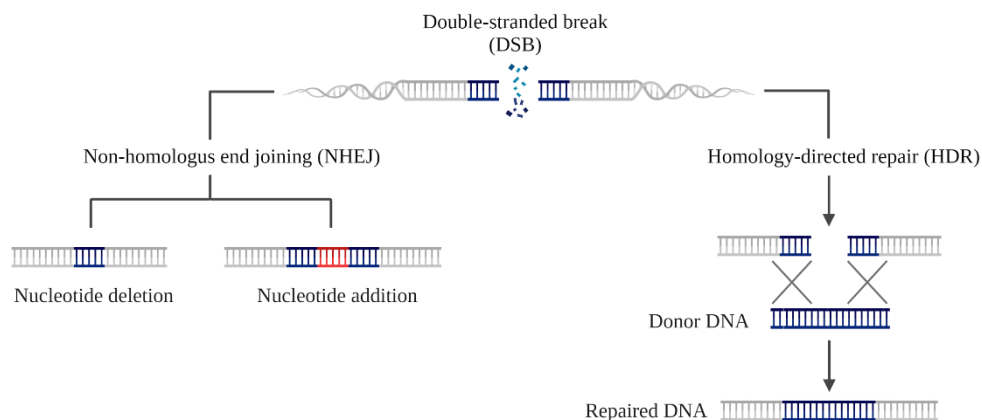


Figure 4. DNA repair pathways in eukaryotic cells to correct double-strand breaks (DSBs): non-homologous end joining (NHEJ, left) and homology-directed repair (HDR, right). Image created with Biorender.

1.3.1. The origins of genome editing tools: from meganucleases to ZFNs and TALENs.

The first genome editing tools available to alter the sequence of specific genes were based on the use of DNA binding proteins^{11,49,52}. The first DNA binding proteins to be applied for genome engineering were the meganucleases. Meganucleases are present in many species, and they are characterized by their capability of binding and excising specific DNA sequences, which can be rather large (Figure 5). They are very specific, but require protein engineering to modify the natural occurring recognition sites, which complicates the experimental design and development^{49,52}.

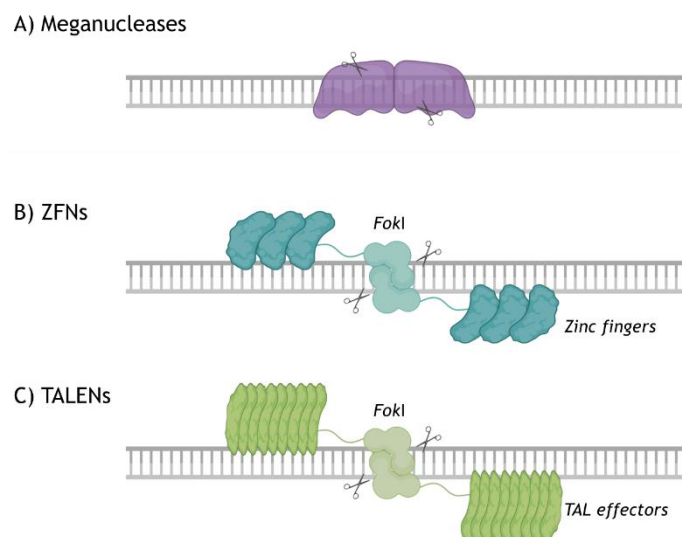


Figure 5. Genome editing systems based on protein-DNA recognition. A) Meganucleases. B) ZFNs. C) TALENs. Image created with Biorender.

The need to find an alternative method to meganucleases, that was easier to adapt lead to the development of ZFN technology in 1996⁵³. ZFN were artificially designed by binding a

DNA cleaving domain from a restriction endonuclease (*FokI*) and a DNA binding domain, which is composed by protein modules that specifically recognize nucleotide triplets. To function, two dimers are needed, one for each DNA strand to be targeted, as indicated in Figure 5. In comparison with meganucleases, ZFN design is simpler, but still challenging to be applied by many laboratories due to both technical and time-consuming issues. However, it has been successfully used to edit the genomes of multiple species both *in vitro* and *in vivo*^{11,49,52}.

The next advance in the field of genome editing occurred in 2010, with the development of a new method based on protein-DNA recognition: TALENs technology⁵⁴. Much like ZFNs, TALENs also rely on the cleaving activity of *FokI*, but they base their recognition capability on transcription activator-like effectors (TALEs) that bind individual base pairs (bp) (Figure 5). This makes them easier to design and synthesize compared to ZFNs, but a protein engineering background is still needed to apply them, which limits their use^{11,49,52}.

1.3.2. CRISPR-Cas9: from prokaryotic immune systems to the new age of genome editing.

The next big step in the genome editing field came from a discovery made in the immune systems of prokaryotes, which was first described in the early 2000s. Many bacteria and archaea have an adaptive defence mechanism against viruses based on CRISPR-Cas. Briefly, in the genome of these organisms there are several *cas* (CRISPR-associated) genes that surround a CRISPR locus (Clustered Regularly Interspaced Short Palindromic Repeats), composed of short direct repeats separated by short variable DNA sequences, also known as spacers. When a virus or plasmid enters for the first time the cell, the exogenous DNA is fragmented and integrated into the CRISPR locus, which originates the spacers. If it's the second time the invasion takes place, then the CRISPR locus it's transcribed along with the *cas* genes. As a result, the exogenous DNA is recognised and cleaved^{55,56}.

Doudna and Charpentier, along with other researchers, worked in the adaptation of this system for genome editing purposes, with the result of the development of the CRISPR-Cas9 system (Figure 6A). For it to work, it needs two components: the Cas9 nuclease and a single guide RNA sequence (sgRNA) containing the tracrRNA and the crRNA regions. The tracrRNA region allows the sgRNA to recognize the Cas9, while the crRNA region binds to a specific DNA sequence by complementarity. Finally, a protospacer adjacent motif (PAM) must be present downstream the DNA targeted sequence. The PAM is a short sequence of 3 nucleotides (5'-NGG-3') that is required for the Cas9 to act⁵⁷.

This original system can be updated with the discovery of new Cas nucleases, with novel PAM sequences and properties, like Cas13 which can target RNA⁵⁸, but also with the application of protein engineering to the already available ones. Cas9 proteins with improved fidelity have been developed⁵⁹, as well as Cas9 versions with non-functional endonuclease domains: one in the Cas9 nickases (nCas9) and two in the dead Cas9 (dCas9). Despite not being able to generate DSBs, these modified versions of the Cas9 can be fused to other protein domains, such as transcriptional repressors (CRISPRi, Figure 6B) or activators (CRISPRa, Figure 6C), which can be applied to regulate gene expression to demand^{60,61}. These engineered Cas9 proteins are also the base of new editing technologies based on CRISPR-Cas9, like base (Figure 6D-E) and prime editing (Figure 6F), that promise to be more efficient and precise, making them more adequate for therapeutic applications^{62,63}.

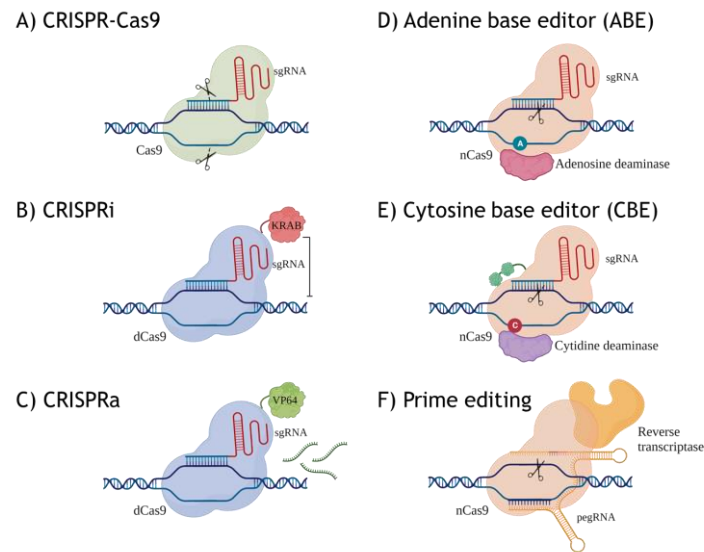


Figure 6. Genomic editing systems based on CRISPR-Cas9. A) CRISPR-Cas9 systems; B) CRISPR interference (CRISPRi) for the repression of gene expression. C) CRISPR activator (CRISPRa) for the enhanced expression of genes. D) Adenine base editor (ABE) for the conversion of adenines into inosines. E) Cytosine base editor (CBE) for the conversion of cytosines into uridine. F) Prime editing system. Image created with Biorender.

1.3.2.1. CRISPR-based systems in biomedicine.

The development of CRISPR-based genomic editing systems has allowed researchers worldwide to develop models to study diseases and disorders with a genetic basis in a simple and affordable way. However, there are still some concerns to the broad application of these technologies for gene and cell therapy approaches, mainly due to the following reasons:

- Potential off-target effects. The sgRNA has the potential to bind to un-targeted regions of the genome with several mismatches, which could generate unwanted edition processes. Despite reports that suggest that the frequency of off-target edits is similar to the typical mutation rate of cells^{64,65}, but there is not a consensus in the field about these findings. This problem could be addressed by optimizing the guide design and performing NGS to identify the presence of off-target genome editing events on the edited cell populations. Additionally, novel systems based on the use of modified nucleases (such as base and prime editing) have a higher fidelity than traditional CRISPR-Cas9, thus reducing the probability of creating off-target modifications in the genome^{66–71}.
- Genome rearrangements due to DSBs. Some authors have reported complex genome rearrangements produced at the on-target and off-target sites as a consequence of the DSB generated by CRISPR-Cas9^{72–74}. Long-read sequencing might help identifying these aberrations that escape traditional validation through Sanger sequencing⁷².
- Immune responses to Cas proteins. The most popular Cas protein is derived from *Streptococcus pyogenes*, a species in close contact with human populations. Therefore, many individuals could have developed an adaptive immune response to their proteins, including Cas9^{75,76}. To try to avoid this issue, many efforts are being dedicated to the identification of Cas proteins from other prokaryotes and to the modification of immunogenic epitopes using protein engineering^{66,77}.
- Toxicity induced by DSBs. It is reported that in some cell lines, such as stem cells, DSBs induce a toxic response mediated by p53, thus reducing the overall efficiency of the process.

Some authors have suggested to transiently inhibit p53 as a way to increase the efficiency of genome editing^{78,79}.

- Delivery of the gene editing system. Gene-therapy delivery is mostly conducted by adeno-associated virus (AAV), but they have a packaging capacity of less than 5 kb which is not enough for the delivery of the large SpCas9 gene (4.1 kb) along with the guide and the other viral components in just one vector. To address this issue some authors have tried to divide the required CRISPR-Cas9 components in two AAV vectors, but this approach has reduced efficiency^{80,81}. Other groups have been working towards the substitution of the large SpCas9 protein with smaller orthologous proteins^{66,77,80}. In addition to this, there are also researchers working on the use of nanoparticles for the delivery of the CRISPR-based systems *in vitro* e *in vivo*^{82,83}.

1.4. STEM CELLS AS AN *IN VITRO* MODEL.

Stem cells are non-specialized cells that have the ability of self-renewal and differentiation into other specialized cell types of an organism. These unique properties make stem cells a very valuable model to decipher developmental mechanisms *in vitro* in an accessible way, which is especially interesting in mammals since most part of this process takes place before birth. In addition, stem cells-based models can be generated for studying for multiple diseases and disorders, and to design new therapeutical approaches based on personalized medicine.

1.4.1. Types of stem cells.

Stem cells can be classified according to their differentiation potential, and to their origin. The first classification can be found in Table 4, ordered from the most to the least potency for the generation of different cell lineages⁸⁴.

Table 4. Stem cell classification attending to their differentiation potential.

TYPE OF STEM CELL	CHARACTERISTICS
Totipotent	They can be differentiated into any cell of an organism, including embryonic and extra-embryonic structures. Example: zygote.
Pluripotent	They can form cells of all germ layers, but not extra-embryonic structures. Example: embryonic stem cells (ESCs).
Multipotent	They can differentiate into cells from a specific cell lineage. Example: neural stem cells.
Oligopotent	They can differentiate into several cell types. Example: lymphocytic stem cells.
Unipotent	They can generate just one cell type. Example: spermatogonial stem cells.

Attending to their origin, stem cells can be classified into somatic stem cells, embryonic stem cells or induced pluripotent stem cells⁸⁴. Somatic stem cells, also known as adult stem cells, can be found within adult tissues and their main function is to renew them. In general, human somatic stem cells are rare and their potency is quite restricted, which has made it difficult to isolate them and use them for research purposes.

Embryonic stem cells (ESCs) derive from the inner cell mass of the blastocyst (4th to 7th day after fertilization), meaning that they are pluripotent. They can be isolated and cultured in the laboratory, but this requires the destruction of the embryo from which they proceed. Most of current knowledge about ESCs derives from studies in mice, due to the ethical concerns related to the use of human ESCs (hESCs)^{84,85}. In Spain, the use of hESCs for research and therapeutical purposes is regulated by the Law of Biomedical Research (*Ley 14/2007, de 3 de julio, de Investigación Biomédica*). In this law, it is stated that the obtention of hESCs for

therapy and research purposes is allowed under certain circumstances. The creation of human embryos for research is prohibited in the country, but hESCs can be obtained from spare pre-embryos from assisted reproduction that were donated for research and therapy purposes. Additionally, all the projects that require the use, obtention, and/or import of hESCs lines need to be review and approved by an specific committee known as *Comisión de Garantías para la Donación y Utilización de Células y Tejidos Humano (Real Decreto 1527/2010, de 15 de noviembre, por el que se regulan la Comisión de Garantías para la Donación y Utilización de Células y Tejidos Humanos y el Registro de Proyectos de Investigación)*^{86–88}.

Induced pluripotent stem cells (iPSCs) were first developed in 2006 by Takahashi and Yamanaka⁸⁹. Briefly, they proved that it was possible to reprogram mouse somatic cells to embryonic-like pluripotent cells by expressing four transcription factors (Pou5f1, Sox2, Klf4, and Myc). A year later, this technique was successfully applied to adult human fibroblasts⁹⁰, which marked a huge step forward in the use of human pluripotent stem cells for research purposes.

1.4.2. hiPSCs as a tool for modelling human diseases and disorders.

In vitro cellular models for the study of human diseases often rely on the use of immortalized cell lines. However, these cell lines often have mutations to allow their long-time culture, but which may have functional consequences as well.

Primary cell cultures from human patients are another interesting source of *in vitro* models, but in general they cannot be maintained in culture for extended periods of time, which is an important limitation for the application of genome editing technologies and long-term experiments. This is particularly relevant for the study of neurological conditions since most neurons are post-mitotic. For this reason, the conduction of extremely invasive brain biopsies to obtain primary human cell cultures for research is unethical^{91,92}.

The optimization of reprogramming technologies to efficiently generate high-quality hiPSCs opened the door to the generation of self-renewing cellular models from a specific individual that preserve their genomic background. Also, when combined with genome editing technologies, it is possible to introduce specific genomic variants to study their functional implications in different cell types. In the following sections, the generation and characterization of hiPSCs will be addressed, as well as their use as *in vitro* models, paying special attention to their relevance for the study of neurodevelopmental disorders.

1.4.2.1. Reprogramming of somatic cells into iPSCs.

As previously mentioned, hiPSCs can be obtained from somatic cells in a process known as reprogramming. At first, adult fibroblasts were reprogrammed by retroviral transduction of four transcription factors (POU5F1, SOX2, KLF4 and MYC), commonly known as the Yamanaka or OSKM factors^{89,90}. Other cocktails of transcription factors have also been successfully used for the same purpose, like the OSNL factors (POU5F1, SOX2, NANOG, and LIN28) reported by Thomson's group⁹³. Also, several somatic cell lines beyond fibroblasts have been successfully reprogrammed, such as peripheral blood mononuclear cells (PBMCs), pancreatic cells, hepatocytes, and adipocytes, among others⁹⁴.

Through the years, novel delivery methods for the reprogramming machinery have been tested, which can be found in Table 5^{94–97}. These methods can be divided into two main groups: integrative and non-integrative delivery. Integrative delivery systems require the insertion of the reprogramming factors into the host's genome, whereas this is not necessary in the case of

non-integrative delivery systems. In general, integrative delivery systems have a higher efficiency, but this comes with a higher risk of insertional mutagenesis, which is not safe for therapeutic applications^{94,96}. The selection of the delivery method strongly depends on the type of somatic cell to be reprogrammed, the efficiency, and the downstream applications of the generated iPSCs.

Table 5. Delivery systems tested for reprogramming of somatic cells into iPSCs.

INTEGRATIVE METHODS	NON-INTEGRATIVE METHODS
Retrovirus	Adenovirus
Lentivirus	Sendai virus
Transposons (piggybac or sleeping beauty)	Episomal vectors
	Protein delivery
	RNA transfection

Nonetheless, the reprogramming efficiency of human somatic cells is still extremely low (0.1 %) and the molecular pathways involved are yet not well understood^{94,95}.

1.4.2.1.1. Characterization of hiPSCs.

After reprogramming, it is extremely important to characterize the obtained hiPSCs to assess their quality.

First, the morphology of the cells must be flat, and they must grow in tightly packaged colonies with sharp edges. Also, good quality hiPSCs must display a normal karyotype and maintain the identity of the original, which is verified using short-tandem repeats (STRs). Additionally, they must be absent of residual reprogramming vectors and have increased levels of the pluripotency markers (POU5F1, NANOG, TRA-1-81, TRA-1-60, etc.). Finally, they also must be able to be differentiated into cells from the three germ layers, which can be assessed by conducting a teratoma or an embryoid body formation assay^{84,94,98}.

When working with hiPSCs, it is recommended to check that these characteristics are still present from time to time under normal culture conditions, but it is even more important if they have been subjected to genome editing.

1.4.2.2. Application of genome editing technologies to disease modelling in hiPSCs.

In general, there are two major strategies for the generation of disease models in hiPSCs. The first one consists in the repair of a potential pathogenic allele in an hiPSCs line derived from a patient, while the second one requires the introduction of potentially pathogenic mutations in healthy hiPSCs. Both strategies end up generating isogenic models that can be compared to identify the presence or absence of disease-related phenotypes^{97,99}.

It is also important to consider the genome editing strategy to be applied: ZFNs, TALENs or CRISPR-Cas9 have been used to edit hiPSCs, the latter has been used the most in recent years due to the technical advantages previously addressed. Several methods have proven to be useful for the successful delivery of the CRISPR-Cas9 components in hiPSCs, such as plasmids, ribonucleoprotein complexes or inducible transgenes. In the end, the choice of the delivery method is deeply influenced by the objective of the experiment (knock-out, knock-in, upregulation, downregulation, etc.) and the efficiency observed in the hiPSCs line of interest^{51,84,99}.

1.4.2.3. Human iPSCs in the study of neurodevelopmental disorders.

The study of human neurodevelopment and its alterations faced the great obstacle of finding a suitable model to do so. Since human primary cultures from brain cannot be obtained from biopsies, for a long time these studies relied on the analysis of human post-mortem samples and the use of animal models^{84,97,100,101}. However, there is a clear difference in brain function between humans and other species, which complicates the study of certain human phenotypes.

This scenario changed with the generation of hiPSCs, which can be used to develop models to study the mechanisms of health and disease in the human brain. Since hiPSCs can be obtained from adult tissue from any patient, this can be applied to study both common and rare diseases to unravel the molecular pathways involved in the disease development and to identify potential therapeutic strategies to revert it^{84,102}.

To date, there are multiple protocols that allow the differentiation of hiPSCs into a large variety of cell types from the CNS (both in 2D and 3D) and many more are under development.

1.4.2.3.1. 2D differentiation of hiPSCs to ASD-related cell types.

The ability to generate specific neuronal subpopulations from hiPSCs has been very useful for the study of many neurodevelopmental disorders, including ASD¹⁰³. Protocols have been developed to produce subtypes of neurons (cortical, GABAergic, Purkinje, dopaminergic, etc.)^{104–107} and glial cells (astrocytes, oligodendrocytes, and microglia)^{108–110}, on which it is now possible to study the effects of different ASD-risk factors at a cellular level^{111,112}. By doing so, certain genotypes can be correlated with the development of alterations in cell morphology and function, at different time points of the differentiation^{97,103}.

However, these models have some limitations that is important to consider. On one hand, the differentiation protocols are not always robust, and a great variability is observed among batches. Also, it is difficult to achieve high levels of maturation, which can be a problem for the identification of phenotypes that appear on later stages of development^{97,103,113}.

1.4.2.3.2. 3D differentiation of hiPSCs for the study of ASD.

The generation of 3D models of the nervous system started with the generation of neurospheres, which are un-organized structures that contain a combination of cell types. However, the level of complexity of these structures was still very limited¹¹⁴. This scenario changed with the development of organoids.

Organoids have been defined as self-organized, three-dimensional structures derived from pluripotent stem cells (or primary tissue stem cells) *in vitro*, which contain multiple cell types and have a cytoarchitecture and function comparable to an organ. There are two types of neural organoids, guided and unguided, depending on the characteristics of the protocol used for their differentiation. Guided neural organoids are generated using exogenous factors to promote the differentiation towards a region of the nervous system. Unguided neural organoids protocols do not rely on exogenous factors, and the obtained structures are more diverse, having neural and non-neural cell types^{115,116}.

In 2013, Lancaster and collaborators developed the first protocol for the generation of whole-brain organoids¹¹⁷, which represented a ground-breaking advance in the field of neuroscience. Since then, many other laboratories have also developed other protocols for the differentiation of neural organoids and spheroids, which represent more specific regions of the

nervous system and are less complex than organoids^{115,116,118,119}. However, spheroids tend to be less heterogenous and can be used to study in depth specific regions of interest within the nervous system.

Brain organoids and spheroids are perhaps the most accurate *in vitro* human model for the recapitulation of *in vivo* neurodevelopment. In this sense, several groups have already applied them to the study of ASD, unravelling the existence of convergent phenotypes despite the aetiology of the disorder^{120–125}. For example, the imbalance between excitatory and inhibitory neurons in ASD, partially due to increased levels of GABAergic neurons.

These models also have technical limitations to overcome, such as the robustness of the protocols, the limited maturation of the obtained organoids, or the lack of vascularization which affects their growth, although some groups are currently working on this issues^{113,126,127}.

1.5. *Danio rerio* AS AN *IN VIVO* MODEL FOR NEUROSCIENCE.

Danio rerio, commonly known as zebrafish, is a small animal native to Southeast Asia that has been widely used as an aquarium fish and, most recently, as a research *in vivo* model in several fields, including neuroscience, as it will be addressed in the following sections.

1.5.1. Anatomical and ecological features of *Danio rerio*.

Danio rerio is a tropical fresh-water fish characterized by a fusiform shape with lateral blue and silver stripes, that naturally inhabits slow streams and pools alongside them formed during the monsoon season. *Danio rerio* is a robust species that can tolerate a wide range of water temperatures (12 - 39 °C), pHs (5.9 - 9.8) and salinities (0.01 - 0.8)^{128,129}.

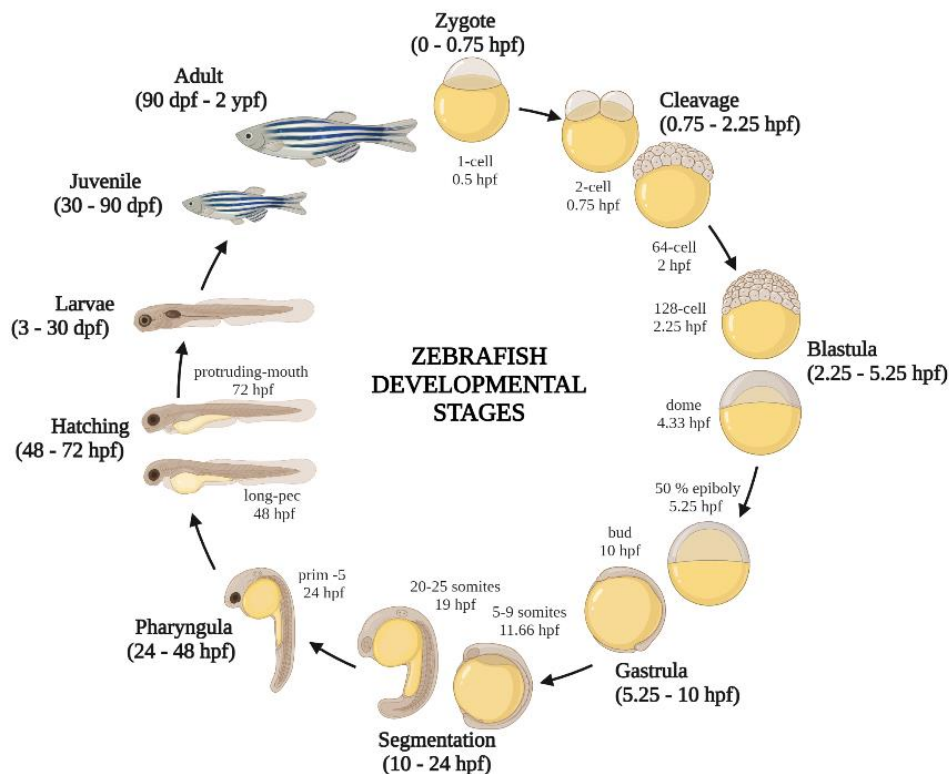


Figure 7. Zebrafish developmental stages. Embryogenesis phases: zygote, cleavage, blastula, gastrula, segmentation, pharyngula and hatching. Image created with Biorender.

Danio rerio is a social fish that lives in shoals, which could be a mechanism for improving survival and reproduction in the wild^{130–133}. Also, these animals display stereotypic mating behaviours, such as the male chasing and touching the female^{134,135}. The domesticated lines of this species lacks heteromorphic sex chromosomes, as well as a defined locus that determines sex, with recent studies pointing that this might be a polygenic character influenced by the environment^{136–138}. For this reason, the identification of males and females mainly relies on sexually dysmorphic features, being the identification of a prominent genital papilla on mature females near the anal fin the most reliable one^{139,140}.

Under laboratory conditions, zebrafish can live for up to several years and reach reproductive maturity in approximately three months, but it is unclear if this is maintained in the wild^{129,141}. As represented in Figure 7, zebrafish embryogenesis is composed of seven phases (zygote, cleavage, blastula, gastrula, segmentation, pharyngula and hatching), that take place on the first three days after fertilization (dpf) and can be defined by the morphological features of the individuals^{142,143}.

1.5.2. *Danio rerio* in biomedical research.

In the 80s decade, *Danio rerio* started to be used as a model organism in areas such as genetics and developmental biology, but recently its use has expanded to other areas as well. This expansion can be explained by the numerous advantages of this model in comparison to other vertebrates^{128,144–148}.

First, the small size, robustness and simple handling of the species make it easy to maintain in a laboratory without a huge investment in comparison with other animal species. Next, zebrafish adults (individuals older than 3 months) can breed on a weekly basis generating around 200 - 300 eggs each time, which is a desirable characteristic for high-throughput experiments. Another interesting characteristic associated with zebrafish's eggs is their transparency and the fact that they are generated by external fecundation. All of this simplify the manipulation and analysis of early developmental changes without the need of using invasive techniques^{128,133,144,147–149}. Other relevant characteristics of zebrafish include the delayed maturation of the immune system, which allows to perform xenografts without the need for immunosuppression, its regeneration capability, and the permeability of the blood-brain-barrier during the first 8 dpf¹⁵⁰.

The research conducted using *Danio rerio* as a model was favoured by the zebrafish genome-sequencing project with onset in 2001, although twelve years were needed for the first high-quality reference genome to be published¹⁵¹. The comparison of zebrafish and human genomes showed that around 70 % of human genes have at least one orthologous sequence in zebrafish, figure that increases to 80 % if only genes associated with human diseases and disorders are considered^{133,148,150,151}.

1.5.2.1. *Danio rerio* in the study of neurodevelopmental disorders.

In the previous section, the characteristics that make zebrafish an interesting *in vivo* model for biomedicine were mentioned. Here, the focus will be put on other traits of this organism that are specifically useful in neurobiology-related studies, for instance, the generation of animal models to study neurodevelopment and its alterations^{133,148–150,152}.

From a structural point of view, zebrafish's CNS has a vertebrate organization, being divided into forebrain, midbrain, hindbrain, and spinal cord^{148,149}. Despite the differences between human and zebrafish CNS development, mainly concerning the neurulation

mechanism and the timeline, in both cases the neural tube derives from ectoderm and the cells organize in an overall similar manner. Mature zebrafish and human brains also have many analogous regions, including the cerebellum, cortex, hippocampus, amygdala, and basal ganglia^{133,148–150,153}. The hypothalamic-pituitary-interrenal (HPI) axis in zebrafish is also analogous of the hypothalamus-pituitary-adrenal (HPA) axis in humans, and both species release cortisol under stress^{148,150,154}.

With regards of neuronal cell types, in zebrafish can be found both neurons and glial cells, including astrocytes, oligodendrocytes, Schwann cells, and microglia^{148,149,153}. Also, human and zebrafish share the main neurotransmitters (including GABA, glutamate, dopamine, adrenaline, serotonin, histamine, and acetylcholine) and some neural circuits associated with the control of basic responses, like sleep and acoustic startle, among others^{133,148,149}.

Finally, *D. rerio* is a social animal that lives in hierarchical shoals and displays a clear preference for conspecifics, a behaviour that is established during the first month of development. Alterations in socialization have been observed and quantified in many zebrafish models for neurodevelopmental disorders, including ASD^{133,150,153}.

1.5.3. Generation of zebrafish models of interest in biomedical research.

External fecundation and transparent embryos are two main characteristics of zebrafish that simplify genomic modification in this species compared to other animals. In this section, the forward and reverse genetics approaches have been applied in zebrafish to study gene function will be discussed^{155,156}.

1.5.3.1. Transient silencing of genes with morpholinos.

Morpholinos (MOs) are synthetic oligonucleotides that can bind a specific mRNA sequence and block its expression (Figure 8). MOs are composed of 18-25 subunits, each one formed by a morpholino ring bound to one of the four nucleotide DNA bases (Figure 9)^{157,158}.

MOs were the first reverse genetics tool used in zebrafish for studying gene function, as its action generates a transient silencing of a specific gene (knockdown). Depending on their mechanism of action, they can be classified into translation or splice blocking MOs, which behave slightly differently in cells. The characteristics of the different types of MOs are collected in Table 6, along with the suitable techniques to verify their action.

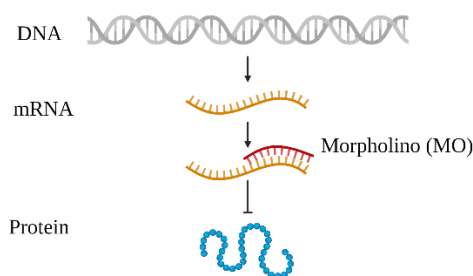


Figure 8. Gene expression silencing (knockdown) mediated by morpholinos. Image created with Biorender.

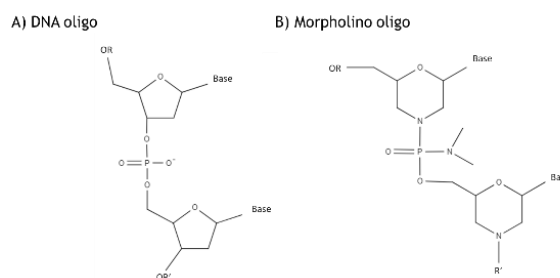


Figure 9. DNA oligonucleotide structure (A) compared to morpholinos (B). Image created with Biorender.

Studies conducted with MOs have several limitations that need to be addressed. On one side, MOs silencing is transient and can only last for about 4 to 5 days, which means that the technique is only useful for short-term experiments. On the other side, MOs can also bind to non-specific regions, which affects the phenotype observed. Also, the morpholine injection can cause the activation of p53-mediated apoptosis, which would also be reflected in the morphant phenotype. To avoid this, it is recommended to co-inject with MOs that target p53, to be certain that the morphant phenotype is not caused by apoptosis and perform rescue experiments with mRNA. Finally, whenever possible, it is interesting to confirm the results in stable mutant models in which long-term analysis can be performed^{148,158–160}.

Table 6. Classification and characteristics of morpholinos.

TYPE OF MO	MECHANISM OF ACTION
Translation blocking	Block the translation initiation complex activity. Do not cause the degradation of kidnapped mRNA.
Splice blocking	Block splicing sites in pre-mRNAs.

1.5.3.2. TILLING.

Targeted Induced Local Lesions in Genomes (TILLING) is a reverse genetics approach that has been widely used to generate mutant zebrafish lines. With this technique, random mutations are introduced in the genome due to exposition to N-ethyl-N-nitrosourea (ENU), which is a mutagen alkylating agent. After exposure, a screening is conducted to identify phenotypes of interest, which can be correlated to ENU-induced mutations. However, this technique requires expensive and time-consuming screenings to identify mutations in specific genes, which is an important limitation to its use^{147,148,161,162}.

1.5.3.3. Advanced gene targeting in zebrafish: ZFNs, TALENs and CRISPR-Cas9.

The development of targeted genome editing technologies facilitated the generation of novel zebrafish mutant lines in an easy and affordable way. Early on, ZFNs and TALENs were used to generate several zebrafish models of interest in neuroscience, but due to their limitations they were quickly replaced due to the development of CRISPR-Cas9 in 2013^{147,148,156,163}.

In the zebrafish field, the standard genome editing with CRISPR-Cas9 requires the *in vitro* transcription of sgRNA and the microinjection of the ribonucleoprotein complex (Cas9/sgRNA) in one-cell embryos. However, this process often generates a population with mosaicism (F0 generation), so it is necessary to identify the individuals with germline mutations, that can transmit them to the next generation. By following this approach, homozygous mutants are obtained after two generations (F2 generation), which usually takes over 6 months^{156,163–167}. To address this limitation, some groups have been working on optimizing the CRISPR-Cas9

methodology to generate full knockouts in the F0 generation using several guides to target the same gene. The F0 zebrafish mutants generated by this approach designated in the literature as crispant mutants to differentiate them from the homozygous mutant generated by the traditional approach^{168,169}. This far, the results of experiments conducted using crispant zebrafish have been very promising, as the individuals recapitulate the full knockout phenotype. While the crispant phenotype might be milder than the observed in homozygous KO mutant lines, crispant phenotypic analysis reduces the amount of time and resources needed to perform the characterization of zebrafish models, which is desirable for high-throughput applications^{168,170–174}.

1.6. FUNCTIONAL CHARACTERIZATION OF *IN VITRO* AND *IN VIVO* ASD MODELS.

In the context of ASD research, it is essential to characterize the generated *in vitro* and *in vivo* models to study the underlying molecular mechanism leading to the onset of the disorder. This section of the introduction is dedicated to present the different methodologies applied in the present thesis to the functional characterization of the generated *in vitro* and *in vivo* models.

1.6.1. Characterization of neuronal differentiation, morphology, and function on *in vitro* models.

With the objective of studying the functional role of potential ASD-risk genes, in the present thesis hiPSCs were differentiated to 2D neuronal cultures and brain organoids. The functional characterization of those models can provide relevant information on their cellular composition and function, thus shedding light onto the mechanisms leading to neurodevelopment and ASD's onset.

First, morphological analysis can be applied to characterize both 2D and 3D hiPSCs derived cultures. These analyses have the objective of assessing neuronal morphology and identifying alterations in neurite outgrowth and synaptic connections. Also, in 3D cultures, the size and shape of the organoids can be studied, which can be relevant for certain ASD cases associated with macrocephaly^{119,123,175,176}.

Also, neuronal marker staining can be performed to verify the successful differentiation of hiPSCs, and the different subtypes of cells present in the cultures. When applying immunofluorescence techniques to identify them (immunocytochemistry or immunohistochemistry) it is also possible to visualize their localization and distribution^{105,119,123,176}.

Additionally, the functionality of the generated neuronal networks (in 2D or 3D models) can be studied with electrophysiological techniques, such as patch-clamp or multi-electrode arrays (MEA). Calcium imaging can also be used to measure intracellular changes in calcium levels, which is related with neuronal activation and connectivity^{105,113,121,123,176}. Nevertheless, the application of these techniques requires specific equipment and specialized training for designing, conducting, and interpreting the experiments, which is not available in every laboratory.

Finally, -OMICS technologies can be also applied to deeply characterize these models, which will be further commented on the following section.

1.6.2. -OMICS technologies for the study of neurodevelopmental disorders.

The nervous system is composed by a network of multiple cell types that interact to receive, process, and transmit information essential for the correct functioning of the organism. The development and functioning of the nervous system are far from being completely understood, and so are the diseases and disorders related to them. In this intricate landscape, high-throughput -OMICS technologies have emerged as a powerful tool for improving our knowledge about neurodevelopment and its alterations^{177,178}. However, most -OMICS technologies are still under development and the number of studies conducted with them is very limited to date, with the exception of genomics, transcriptomics, and proteomics. The relevance of genomics approaches to study the genetic variability associated with ASD has already been commented on in this introduction (Section 1.2.2.2.1), so this section will be dedicated to the relevance of novel transcriptomic and proteomic approaches to study ASD.

1.6.2.1. Application of single-cell RNA sequencing to ASD research.

Traditional RNA sequencing (RNA-seq) techniques relied on the transcriptomic analysis of RNA samples derived from mixed cellular populations. Nevertheless, this analysis renders the average gene expression change of the population, masking the heterogeneity of the original sample. In 2009, this limitation was addressed by the development of the first single-cell RNA sequencing (scRNA-seq) techniques, which allow to analyse the expression patterns of individual cells¹⁷⁹.

By using scRNA-seq techniques it is now possible to precisely identify cell populations based on their expression patterns, as well as identifying minor cell types within complex samples^{179,180}. Technically, five steps are needed to conduct a scRNA-seq experiment, which are the following¹⁷⁹:

- Isolation of single cells. There are different techniques to perform the isolation, depending on the sample of interest.
- Reverse-transcription. RNA obtained from single cells is reverse-transcribed to complementary DNA (cDNA).
- Library construction. The cDNA is amplified and used for the generation of a representative library.
- Library sequencing. The obtained libraries are sequenced using high-throughput techniques.
- Data analysis. Specific bioinformatic pipelines are applied to identify expression patterns at a single-cell level.

Through the application of scRNA-seq, an enrichment in the expression of ASD-risk genes was found in human inhibitory neurons, which supports the hypothesis of an imbalance between excitatory and inhibitory neurons underlying ASD pathogenesis^{179,181–183}. Additionally, single-cell transcriptomics of brain organoids with alterations in ASD-risk genes also pointed out to the altered formation of GABAergic and deep-layer excitatory projection neurons as a common phenotype in this disorder. Moreover, this technology also helped in the identification of microglia cells are potentially implicated in ASD-pathogenesis, which was impossible to identify using conventional RNA-seq transcriptomics.

Despite being a promising technique, there are still challenges to scRNA-seq application. First, this technology is very expensive in comparison with traditional RNA-seq, which prevents many laboratories from applying it. Second, scRNA-seq requires the isolation of cells,

which can potentially alter their transcriptome. Third, low-abundance transcripts might not be detected. Fourth, the technique is very sensitive to technical variations, which might affect the obtained results. Finally, scRNA-seq renders a huge amount of data, which needs to be properly analysed and interpreted. Notwithstanding, the field is constantly changing to implement improved versions of the protocols and bioinformatical pipelines to overcome these limitations.

All in all, scRNA-seq holds promise in furthering our comprehension of brain function in both healthy and disease states. By harnessing the power of this technology, we can anticipate significant advancements in our understanding of neurological conditions.

1.6.2.2. Application of proteomic analysis to ASD research.

Proteomics consist in the large-scale identification of proteins expressed within a sample, with the potential of identifying distinct protein isoforms and post-translational modifications^{184,185}. Thus, the application of proteomics to the study of human diseases and disorders has a great potential to unravel their molecular mechanisms, as well as to identifying novel biomarkers and therapeutical targets^{185,186}. Nowadays, mass-spectrometry (MS) is the most popular technique for the identification and quantification of proteins within biological samples. Briefly, mass spectrometry experiments require three key steps: sample preparation, data acquisition, and data analysis^{187,188}.

The first step, sample preparation, requires protein extraction and ionization. Proteins can be extracted from multiple biological sources (from cell cultures to complex tissues), and specific purification protocols can be applied to isolate proteins within cellular compartments or harbouring specific modifications, such as phosphorylation. After extraction, proteins can be fragmented into peptides (bottom-up) or analysed directly (top-bottom). Despite losing some peptides in the process, bottom-up has been preferred over top-bottom since it is easier to separate and ionize peptides than proteins^{187,188}.

Next, proteins need to be ionized to be detected by MS. The most common tool for the ionization and analysis of complex samples is liquid chromatography (LC), while matrix-assisted laser desorption ionization (MALDI) is preferred for simple mixtures of proteins/peptides^{187,188}.

Finally, each identified peptide needs to be mapped to their corresponding protein, which is dependent on the quality of the protein's annotations available. Proteomic analysis would not be complete without tools to give biological meaning to the findings, which are often based on the gene ontology (GO) terms associated with the identified proteins¹⁸⁸.

Despite their potential, few proteomic analysis have been conducted in the field of ASD research¹⁸⁴⁻¹⁸⁸. Some studies have tried to identify common proteomic alterations in ASD cases using post-mortem brain samples. However, those samples are not the most suitable to understand the molecular mechanisms underlying ASD, since cell-death undoubtedly affects the results. Other authors have analysed non-neuronal biological samples obtained from ASD-affected individuals, although few conclusions could be obtained from those studies as the affected tissue was not analysed.

Proteomic analyses of human iPSCs-derived neurons and brain organoids are an interesting alternative for advancing in ASD's research. Despite not being able to fully recapitulate the complexity of human neurodevelopment, pre- and post-synaptic alterations associated with ASD have been identified using this novel approach¹⁸⁷.

Overall, continuous advancements in proteomic techniques and hiPSC-based modelling are expected to improve the current understanding of ASD pathophysiology. In addition, novel

biomarkers and potential therapeutical targets could be explored with this novel approach in the near future^{187,188}.

1.6.3. Phenotypic characterization of zebrafish models.

Multiple zebrafish morphant and mutant lines have been used to study ASD-risk genes and the potential mechanisms underlying the disorder. As previously mentioned, these models are of great interest to study the phenotypic consequences of the alteration of certain genes, which is the objective of their use in the present thesis. In this section, morphometric and behavioural characterization of zebrafish models will be addressed.

1.6.3.1. Morphometric characterization.

One of the main advantages of using zebrafish as a model organism is the transparency and accessibility of the developing embryos. These characteristics make possible to conduct phenotypic measurements of the developing embryos without invasive techniques. In the context of ASD research, several body measurements have been widely used as indicators of the neurodevelopmental state of the individuals, as they correlate with alterations in this process. Body length, eye area, head size, and otolith shape are the most common phenotypic measures to characterize the development of zebrafish larvae^{148,189–193}.

1.6.3.2. Behavioural characterization.

As discussed earlier on this introduction, ASD are mainly characterized by the presence of behavioural alterations in individuals. For this reason, it is important to address the presence of alterations in this area when a novel model for ASD is generated. Over the years, researchers have developed multiple tests to analyse behavioural changes in zebrafish, both in adults and larvae. This section provides a summary of the tests most commonly use in ASD research^{150,152,153,194}.

- Social preference. *Danio rerio*'s preference for conspecifics can be easily analysed by tracking the response of an isolated individual to the exposure of other zebrafish in a neighbouring tank. This test also can be used to analyse the preference of the individual for known fish rather than unknown conspecifics^{190,194–196}.
- Shoaling formation. Adult zebrafish display a preference for swimming in shoals rather than alone. This behaviour can be analysed by measuring the distance to the nearest and farthest fish, the amount of time spent in/out the shoal, as well as the presence of a uniform direction of swimming in the group^{133,150,190,194}.
- Inhibitory avoidance. This test is designed to analyse the learning capability of the individuals, which are exposed to attractive or aversive *stimuli*¹⁹⁴.
- Aggression. The presence of aggressive behaviour in zebrafish can be analysed by using the flat mirror test, which consists of quantifying the number of attacks to the mirror¹⁹⁴.
- Repetitive behaviours. Zebrafish models for ASD often display repetitive behaviours, for instance, stereotypic mouth opening and circling swimming^{194,197}.
- Thigmotaxis. The thigmotaxis can be measured by analysing the amount of time spent in the periphery of the water tank, which is correlates with the levels of anxiety of the fish^{190,198,199}.
- Locomotor activity. Zebrafish response to light mechanical *stimuli* or light/dark cycles is recorded and analysed^{190,199–201}.

1.7. APPLICATION OF *IN VITRO* AND *IN VIVO* MODELS OF ASD: FROM THE MOLECULAR MECHANISMS TO PERSONALIZED MEDICINE.

The generation of *in vitro* and *in vivo* models that replicate ASD-associated genetic variants hold significant promise in deciphering the molecular mechanisms underlying the disorder. The knowledge generated as a result of these studies not only deepen our comprehension of the complex process of human neurodevelopment, but also can be applied to improve the quality of life of individuals with ASD.

On one hand, providing evidence of a functional consequence linked to a specific mutation can improve the diagnosis, and clinical and pre-implantation management of the disorder and the genetic counselling provided to the affected individual and their relatives.

Additionally, these models can be used to develop and optimize novel therapies that target the core symptoms of the disorder, which are not yet available. Several groups are focused on the development of small-molecule drugs for ASD treatment, but gene therapy approaches are also being explored. Gene therapy is based on the alteration of gene expression as a mechanism to rescue a pathogenic phenotype, which can be achieved by transgene delivery in the organism or genome editing technologies. However, there are several limitations that both approaches need to overcome before they can be tested on clinical trials. The first one is related to the delivery method, which ideally should be tissue specific and able to penetrate the blood-brain-barrier (BBB). The packaging size and immunoreactivity of the delivery method also needs consideration since the efficacy of the treatment strongly depends on them. Also, it is extremely important to be able to detect potential off-target effects that could occur when genome editing technologies are being applied and develop editing systems free of them^{202–205}.

Overall, *in vivo* and *in vitro* models of human diseases represent an unparalleled biomedical advancement for precision medicine, personalized diagnosis, and prognosis, but also in the tailored development of novel therapies. Thus, they constitute the foundation for the evolution of this new biomedicine paradigm that places the patient at the core of therapeutic progress.

1.8. SELECTION OF GENES TO BE STUDIED IN THIS THESIS.

The present thesis is focused on developing suitable *in vitro* and *in vivo* models to study the functional implications of genomic variants potentially associated with ASD development. This section provides a comprehensive analysis of the selected genes to be modeled in this thesis: *BTBD8*, *FOSL2*, and *TAOK1*. Subsequent subheadings will offer detailed insights into these genes, encompassing their expression patterns, roles in cellular function based on current publications, and their association with the onset of human diseases and disorders. Additionally, this section also explores the feasibility of using *Danio rerio* as an *in vivo* model.

1.8.1. *BTBD8*.

According to GRCh38 genome assembly, the gene BTB Domain Containing 8 (*BTBD8*, ENSG00000189195, HGNC: #21019) can be found in chromosome 1p22.1 (chr1: 92080305-92184725), and it encodes 8 transcripts, 4 of them classified as protein coding²⁰⁶. In Figure 10 can be found the *BTBD8* expression pattern in different tissues and cell types, being expressed mainly in the foetal brain^{207,208}.

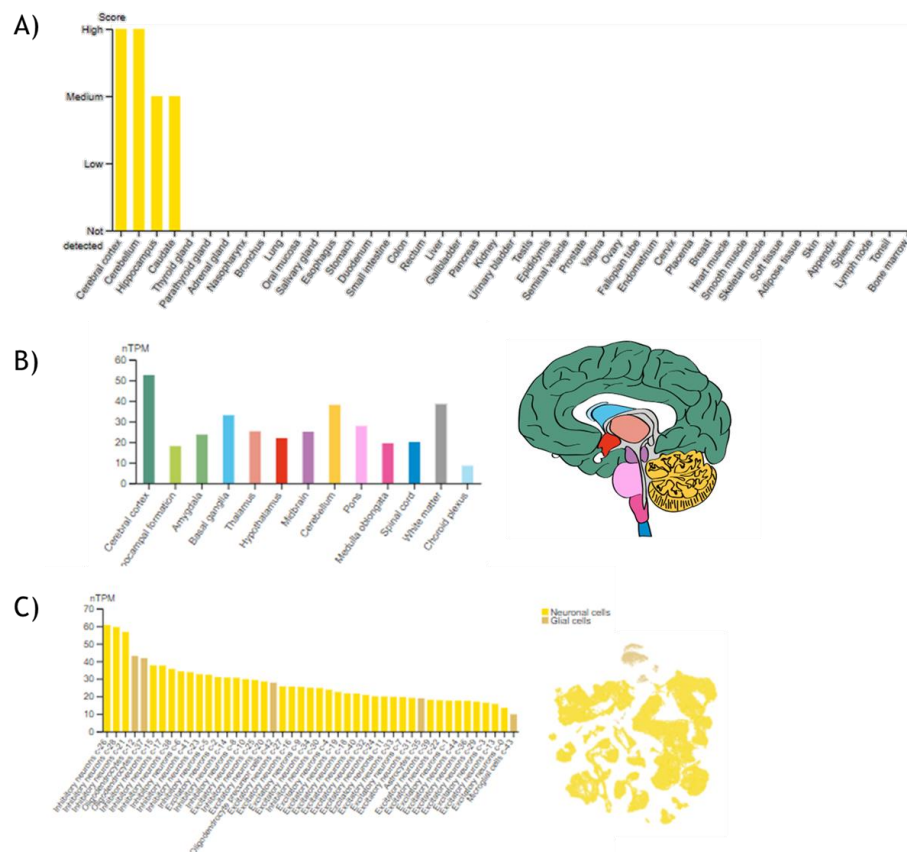


Figure 10. *BTBD8* expression pattern in different tissues and cell types. A) Global *BTBD8* protein expression across different human tissues. B) *BTBD8* RNA expression across different human brain regions. C) *BTBD8* RNA expression at the single cell level in human brain. Expression levels indicated in normalized transcript per million (nTPM) Images obtained from The Human Protein Atlas database (Creative Commons Attribution-ShareAlike 3.0 International License) ²⁰⁷.

The main BTBD8 protein isoform has 1792 aa (Uniprot: #Q5XXL5) and localizes differently in foetal and adult cells, according to published information. In foetal-origin cells, such as HEK293, the protein was found in the nucleus, while in adult cells, like NIH/3T3, it is present in the cytoplasm. BTBD8 lacks nuclear localization signals, so it is likely that its translocation is dependent on other proteins. Also, in primary murine cultures, it was found that BTBD8 localization varies from immature to mature neurons. In the first ones, it is present in the cytoplasm and growing processes, while in mature neurons it localizes mainly in presynaptic terminals and axons^{209–211}.

BTBD8 main isoform, represented in Figure 11, only contains two protein domains, known as “broad-complex, tramtrack, and bric-à-brac/poxvirus and zinc finger” (BTB/POZ), which are evolutionary conserved and often found in transcription factors involved in the regulation of development^{209,210,212,213}.

BTBD8 has been identified as a component of the clathrin-mediated endocytosis at the synapse, process that will be discussed more in detail in the following section.

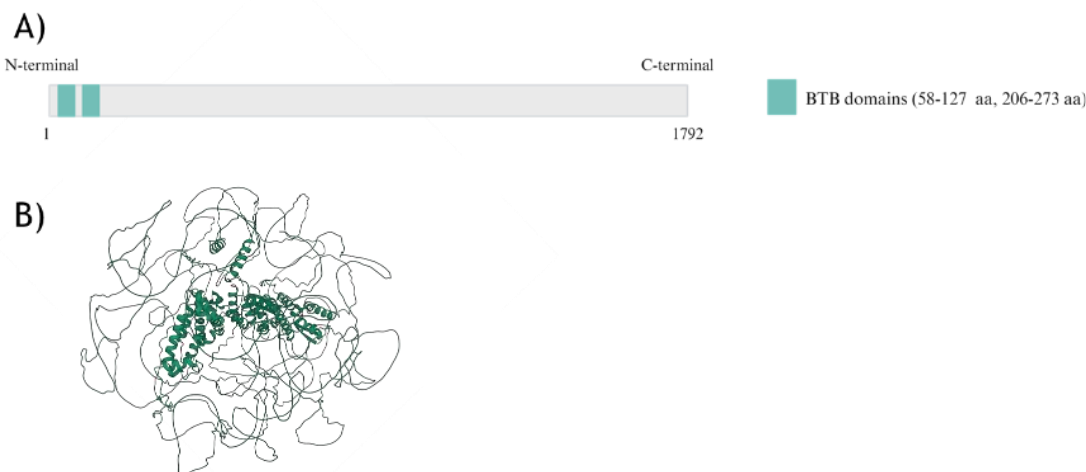


Figure 11. Protein structure of BTBD8 main isoform. A) Schematic representation of the protein domains in BTBD8 (representation created by the author). B) BTBD8 predicted 3D structure by AlphaFold Monomer v2.0 pipeline (Creative Commons Attribution 4.0 license)^{214,215}.

1.8.1.1. BTBD8 role in synaptic function.

At the present time, the available information about BTBD8 function is very limited, with only one article describing its role published by Piccini *et al.* in 2017²¹¹. The authors of this article applied a bioinformatic pipeline to identify uncharacterized genes potentially associated with synaptic vesicle recycling, identifying *BTBD8* (previously known as *KIAA1107*) as a result. Following that finding, the authors conducted *in vitro* assays in SH-SY5Y human neuroblastoma cells and mouse brain extracts to analyse the role of BTBD8 in synaptic function. As a result, they found out that BTBD8 is a part of the clathrin-mediated endocytosis system at the synapse through its interaction with adaptor protein-2 (AP-2). In addition to this, in the same article the authors conducted downregulation assays in mouse primary cultures, that suggest a significant role of BTBD8 in early neurodevelopment and in synaptic transmission.

1.8.1.2. Phenotypes, diseases, and traits associated with BTBD8 variants.

The available information about the role of BTBD8 is extremely limited, and so it is the data concerning its implication in the development of human diseases and disorders.

As mentioned in the previous subheadings, *BTBD8* is thought to play a key role in early development and synaptic transmission^{210,211}, therefore, mutations in this gene could be associated with developmental alterations. Supporting this hypothesis, Alonso-González and collaborators found a germinal mutation in this gene (NP_001363060.1:p.Ser937*) potentially associated with ASD development in a Spanish cohort of 360 patients and their unaffected parents^{26,216}.

Apart from that, only a couple of studies have linked alterations in this gene with other human pathologies. For instance, a missense variant in *BTBD8* (p.Val60Ile) has been associated with a lower monocyte count²¹⁷ and the expression levels of *BTBD8* have been correlated with the prognosis of malignant meningioma²¹⁸.

1.8.1.3. Orthologous genes of *BTBD8* in zebrafish.

Although *BTBD8* has 136 recognized orthologous sequences in multiple species, none has been described in *D. rerio* at the moment²⁰⁶.

1.8.2. *FOSL2*.

FOS-like 2 (*FOSL2*, ENSG00000075426, HGNC:3798) is a member of the FOS family, which also includes *FOS*, *FOSB* and *FOSL1*. These four genes encode leucine-zipper proteins that interact with members of the JUN family, thereby forming a complex called Activator Protein 1 (AP-1) which acts as a transcription factor^{219–221}.

According to GRCh38 genome assembly, *FOSL2* is located in chromosome 2p23.2 (chr2:28392858-28417317) and has three protein coding transcripts formed by alternative splicing, which originate three protein isoforms (202, 318 and 326 aa). The canonical protein isoform, 326 aa (Uniprot: #P15408), is present in the cell nucleus and contains a basic leucine zipper domain (bZIP), that consists of a basic region and a leucine zipper (Figure 12). The predicted function of the bZIP domain is to stabilize dimerization with other bZIP harbouring proteins thanks to the formation of alpha helices^{206,209,212}.

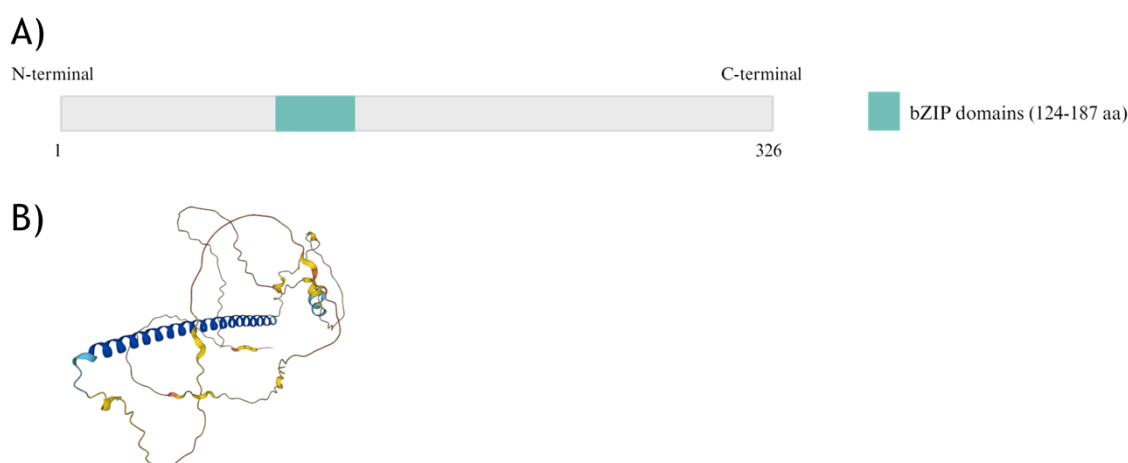


Figure 12. Protein structure of FOSL2 main isoform (representation created by the author). A) Schematic representation of the protein domains in FOSL2. B) FOSL2 predicted 3D structure by AlphaFold Monomer v2.0 pipeline (Creative Commons Attribution 4.0 license)^{214,215}.

FOSL2 expression has low tissue specificity, as it exhibits high levels of protein in many tissues. In Figure 13 can be found a list of organs with detected expression of *FOSL2* ordered by their level of expression, which include the cerebellum, adrenal glands, and organs of the digestive and respiratory systems, among others²⁰⁷.

Members of the FOS family are implicated to the regulation of cell proliferation, migration, survival, and differentiation²²¹.

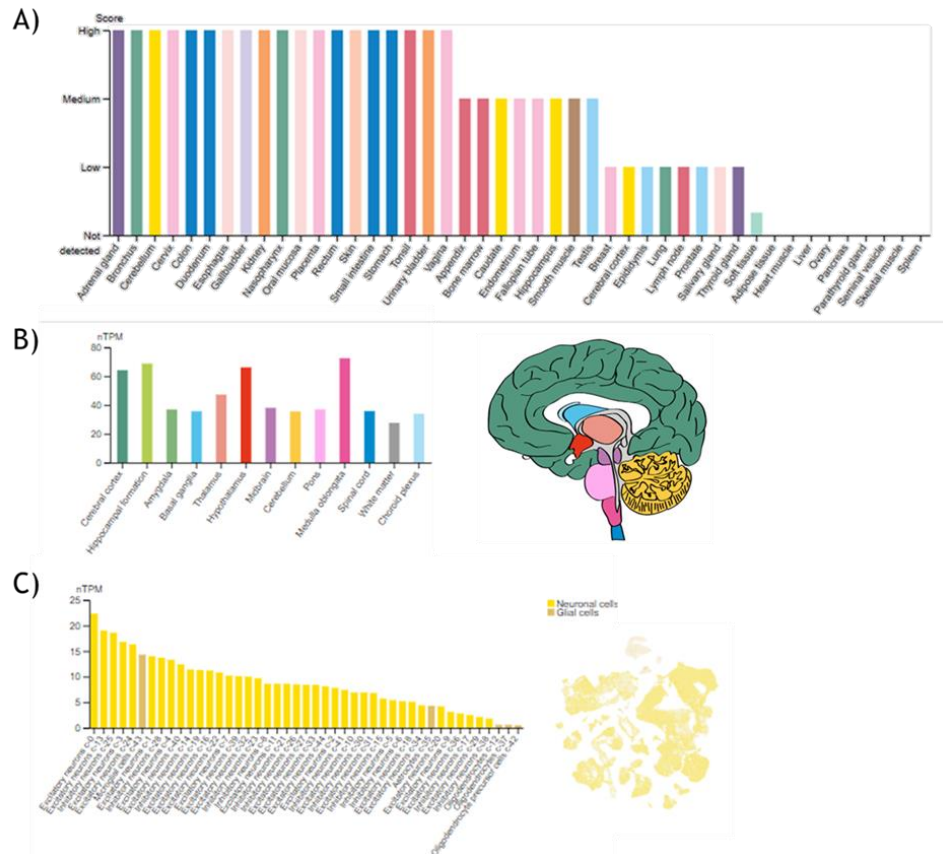


Figure 13. *FOSL2* expression pattern in different tissues and cell types. A) Global *FOSL2* protein expression across different human tissues. B) *FOSL2* RNA expression across different human brain regions. C) *FOSL2* RNA expression at the single cell level in human brain. Expression levels indicated in normalized transcript per million (nTPM) Images obtained from The Human Protein Atlas database (Creative Commons Attribution-ShareAlike 3.0 International License)²⁰⁷.

1.8.2.1. AP-1 transcriptional complex: composition and function.

The AP-1 transcription factor family plays a crucial role in the regulation of cell expression in response to environmental signals. It is comprised of dimeric transcription factors formed by homo- or heterodimerization of members of four protein families: JUN, ATF, MAF, and FOS. However, not all AP-1 combinations are possible. For example, FOS family members are unable to form dimers with members of the same family^{221,222}.

To date, the best studied AP-1 dimers are the ones formed by members of the FOS and JUN families (JUN, JUNB, JUND), which preferentially bind DNA motifs known as 12-O-tetradecanoylphorbol-13-acetate (TPA)-responsive elements²²².

AP-1 transcription factors are widespread across different cell types and tissues. Also, their composition is very dynamic, and it quickly changes in response to external signals, altering the binding capability of the AP-1 transcription complex²²¹. It is known that AP-1 dysregulation can lead to severe diseases, including cancer and chronic inflammatory pathologies²²². However, given the intricate dynamics involved in AP-1 gene regulation, it is plausible that malfunctioning of AP-1 could contribute to the development of numerous other pathological conditions.

1.8.2.2. Phenotypes, diseases, and traits associated with *FOSL2* variants.

Most studies concerning *FOSL2* function have been focus on studying its role in cancer metastasis, since it's a regulator of cell migration and invasion^{223–229}. In fact, some of them have even addressed the possibility of targeting *FOSL2* as a therapeutic approach to treat some cancers^{224,225,227,230,231}.

FOSL2 has also been studied in the context of inflammatory and autoimmune diseases due to its role in the regulation of the differentiation of cells from the immune system, including both B and T lymphocytes^{221,232–235}.

The correct function of this gene is also important in osteoclasts, controlling their size and survival²³⁶, ameloblasts, suggesting a role in enamel formation²³⁷, and cardiomyocytes, regulating their differentiation²³⁸. *FOSL2* role in fibrosis has also been investigated in several tissues, including heart²³⁹ and lung²⁴⁰.

Finally, *FOSL2* alterations have also been associated to neural function and ASD development. Loss-of-function mutations in neuronal ATP-dependent chromatin remodelling subunits, such as *ACTL6B* or *SMARCA2*, which affects chromatin accessibility. The main consequence of this seems to be the abnormal activation of the AP-1 transcription complex^{241,242}, which leads to the development of an ASD phenotype. Related to this, in December 2022, Cospain and collaborators²²¹ published an article describing a direct association between mutations in *FOSL2* and the presence of a syndromic developmental disorder. This disorder seems to have a common phenotype across individuals from different families, characterized by developmental delay, scalp aplasia, enamel abnormalities, congenic cataracts, and ASD. The authors where able to identify the presence of heterozygous truncating mutations in the last exon of *FOSL2* in all the affected individuals. The authors conclude that mutations in this region allow *FOSL2* mRNA to escape non-sense mediated decay and to maintain its interaction with JUN longer than usual. Therefore, the AP-1 complex is stabilized which alters its dynamic function.

1.8.2.3. Orthologous genes of *FOSL2* in zebrafish.

FOSL2 has 220 orthologous sequences, two of them identified in zebrafish. The high confidence orthologous gene in this species is known as *fosl2* (ENSDARG00000040623) and has a 64 % homology with the human sequence²⁰⁶. Localized in chromosome 17 (chr17:41302660-41318547), *fosl2* has only one transcript that encodes a protein of 341 aa²⁰⁶. The expression of this gene is low between the gastrula and the hatching stages, increasing significantly during the larvae phase²⁴³.

The other *FOSL2* orthologous in zebrafish is ENSDARG00000092174, which is localized in chromosome 20 (chr20:9211237-9216772), but in this case, it has a low conservation (< 40 %) with the human sequence²⁰⁶.

1.8.3. *TAOK1*.

The *Thousand And One Kinase 1* (*TAOK1*, ENSG00000160551, HGNC: #29259) encodes a protein that is part of the Ste20-related mitogen-activated protein kinase kinase kinase (MAP3K) family. *TAOK1* is localized in chromosome 17q11.2 (chr17: 29390363-29551903, GRCh38) and it encodes 6 transcripts, 4 of which are classified as protein coding and the remaining two as retained introns^{206,219,244}. *TAOK1* expression is high in several tissues, especially in brain and prostate, as indicated in Figure 14²⁰⁷.

TAOK1 is a member of the thousand and one amino acid kinases (TAOKs) subfamily, which is involved in the regulation of key signalling pathways in cells. TAOKs, as members of the MAP3K family, catalyse the transference of phosphate groups into serine/threonine residues of their substrates, which results in the alteration of their activity^{206,219,245–248}. In mammals, the TAOK subfamily (TAOK1, 2 and 3) of proteins has been identified as regulators of several signalling pathways (MAPK and Salvador–Warts–Hippo pathway), but they are also able to interact with cellular proteins independently of their kinase activity^{245,247}.

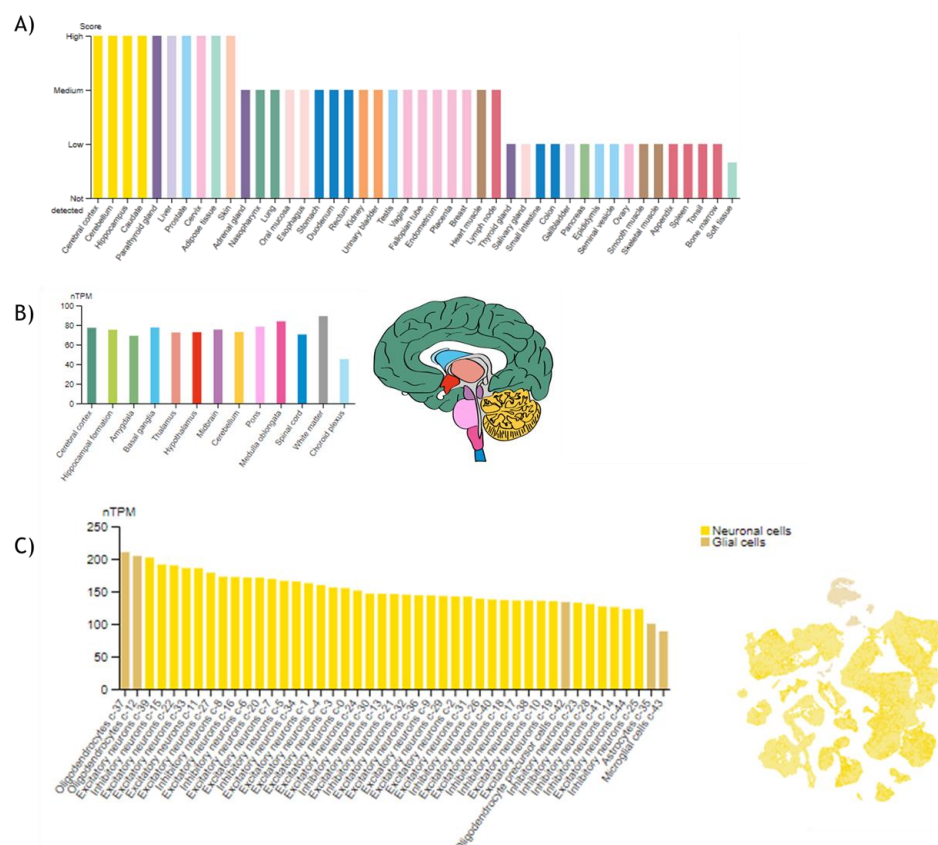


Figure 14. *TAOK1* expression pattern in different tissues and cell types. A) Global *TAOK1* protein expression across different human tissues. B) *TAOK1* RNA expression across different human brain regions. C) *TAOK1* RNA expression at the single cell level in human brain. Expression levels indicated in normalized transcript per million (nTPM). Images obtained from The Human Protein Atlas database (Creative Commons Attribution-ShareAlike 3.0 International License)²⁰⁷.

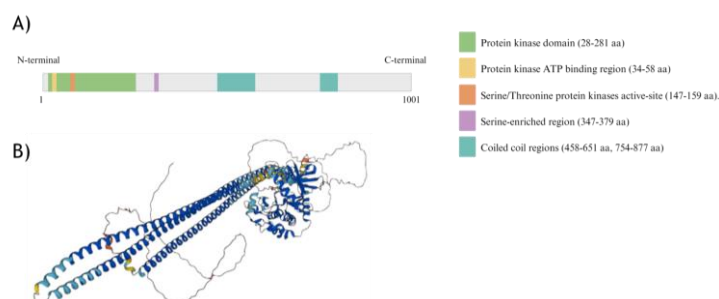


Figure 15. Protein structure of TAOK1 main isoform. A) Schematic representation of the protein domains in TAOK1 (image created by the author). B) TAOK1 predicted 3D structure by AlphaFold Monomer v2.0 pipeline (Creative Commons Attribution 4.0 license)^{214,215}.

As the family's name indicates, the main TAOK1 protein isoform has 1001 amino acids (aa) (Uniprot: #Q7L7X3) and localizes in vesicles the cytoplasm. Figure 15 shows a schematic representation of its structure, featuring in the N-terminal region the protein kinase domain (28-281 aa), including an ATP binding region (34-58 aa) and the predicted active site (147-159 aa). Near the kinase domain, there is a serine enriched region, followed by two coiled-coil domains in the C-terminal^{206,207,209,212,245,247}.

1.8.3.1. TAOK1 role in cellular function.

As mentioned previously, TAOK proteins are part of several signalling pathways that regulate the cellular response to external and internal stimuli^{245,247,249}.

- MAPK pathways: p38/MAPK, and mitogen-activated protein kinase 8 (MAPK8, previously known as JNK) pathways. These pathways are activated by cellular or extracellular stimuli, that trigger the activation of a MAP3K protein (such as TAOKs). Then, this kinase activates a MAP2K that acts on a MAPK. As a result of this cascade, MAPK phosphorylates intracellular targets altering protein function or gene expression in response to the original stimuli.
- Salvador-Warts-Hippo pathway (SWH). This pathway also acts through a series of kinases that act sequentially to finally phosphorylate transcription co-activators YAP1 (previously known as YAP) and TAFAZZIN (previously known as TAZ). By acting upon them, this cascade regulates the proliferation/apoptosis of cells, thus controlling organ size.

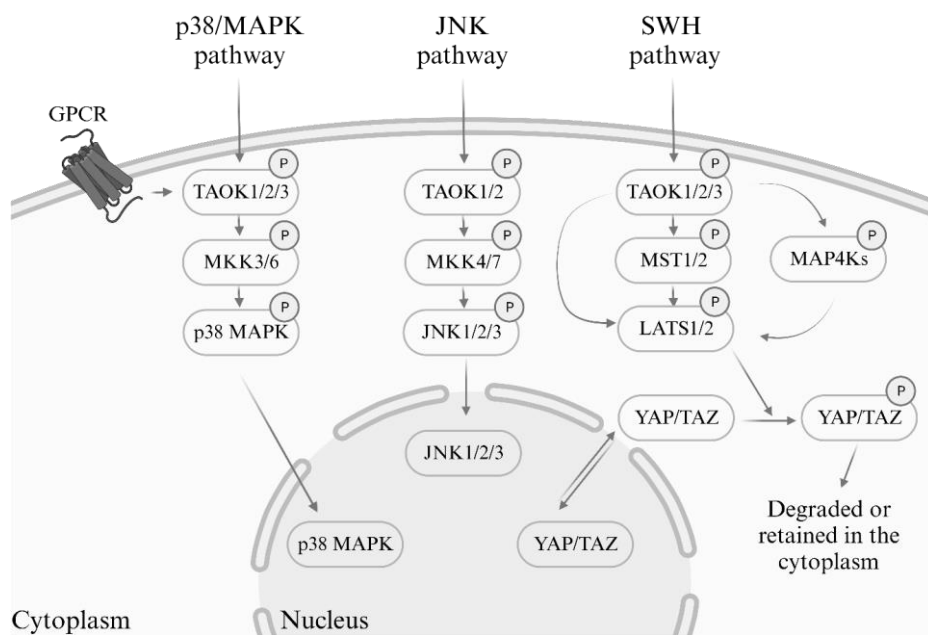


Figure 16. TAOK1 role in the MAPK and Salvador-Warts-Hippo pathways. Image created with Biorender.

TAOK1 and TAOK2 are also involved in the regulation of several cytoskeleton components, such as microtubules and actin, playing an important role during mitosis and neuron formation and migration. Both are needed for spindle positioning and cell rounding during mitosis: TAOK1 phosphorylates microtubule-binding proteins, like tau, inducing microtubule instability, while TAOK2 stabilizes microtubules through direct binding^{245,247,250,251}. In neurons, TAOK1 activates MARK2 (microtubule affinity-regulating kinase), which acts upon tau, allowing microtubule rearrangements responsible for neurite

formation^{248,252–254}. Also, TAOK1 and TAOK2 participate in synapse development by binding to myosin Va, which is possible after being phosphorylated by STK24^{245,247,255}. A recent study by Beeman and collaborators demonstrated that TAOK1 also acts as a plasma membrane remodelling kinase by directly binding to phosphoinositides, an activity regulated by autophosphorylation²⁵⁶.

Members of the TAOK family are also involved in the regulation of immunity, inflammation, and apoptosis. In this sense, TAOK1 inhibits IL-17-mediated inflammation and induces the production of pro-inflammatory cytokines^{247,257–260}. TAOKs also trigger the morphological changes associated to apoptosis, through their activity as cytoskeleton remodelling proteins. Also, apoptotic agents can activate both TAOK1 and TAOK2, which are able to activate apoptotic-related targets including caspases and other kinases such as Rho Kinase 1 (ROCK1)^{245,247,261}.

All in all, the functions of TAOK1 and other members of its family are far from being fully characterized due to the complexity of the molecular pathways in which they participate.

1.8.3.2. Phenotypes, diseases, and traits associated with *TAOK1* variants.

TAOKs disfunction has been linked to cancer development. Although this connection has not been fully investigated, some studies identified that TAOKs knockdown impair the mitotic DNA-damage response, as well as the cell response to apoptotic stimuli, thus leading to tumorigenesis^{245,247,251,262–267}.

TAOKs are also related to neurodegenerative diseases and neurodevelopmental disorders. Related to the first ones, some reports state that TAOK1 and TAOK2 might participate in the abnormal tau phosphorylation that is observed in Alzheimer's disease^{245,247,268}, and TAOK3 in the kinase cascade leading to tau phosphorylation in Parkinson's disease^{245,269}.

Table 7. *TAOK1* mutations associated with developmental delay with or without intellectual impairment or behavioural abnormalities (MIM 619575).

REFERENCE	TYPE OF MUTATION	INHERITANCE
Dulovic-Mahlow <i>et al.</i> , 2019 ²⁵³	Missense variants (n = 4)	<i>De novo</i> (all)
	Nonsense variants (n = 3)	<i>De novo</i> (all)
	Frameshift variants (n = 1)	<i>De novo</i> (all)
Woerden <i>et al.</i> , 2021 ²⁴⁸	Missense variants (n = 5)	Unknown (n=1) <i>De novo</i> (n=4)
	Nonsense variants (n = 7)	Inherited (n=2)
		Unknown (n=2)
		<i>De novo</i> (n=3)
	Frameshift variants (n = 5)	<i>De novo</i> (all)
	Splice site variants (n = 3)	<i>De novo</i> (all)
Hunter <i>et al.</i> , 2022 ²⁴⁶	Missense variants (n = 1)	<i>De novo</i> (all)
	Splice site variants (n = 1)	<i>De novo</i> (all)
	Frameshift variants (n = 2)	Inherited (all)
Blaschek <i>et al.</i> , 2023 ²⁷³	Frameshift variants (n = 1)	<i>De novo</i> (all)

Related to the TAOKs association with neurodevelopmental disorders, to date, *TAOK1* is identified as a syndromic, high confidence ASD-associated gene in the SFARI database³². It was first identified as a potential ASD-associated gene, as a result of exome-sequencing studies conducted in cohorts of ASD patients^{27,42,216}. Since then, additional reports have been published where mutations in this gene were identified as the cause of a developmental delay with or without intellectual impairment or behavioural abnormalities (MIM 619575). Despite being quite variable, several studies identified some common symptoms of this disorder which are developmental delay, intellectual disability, feeding difficulties, hypotonia, behavioural

problems, joint laxity and craniofacial dysmorphism. It is important to notice that only 33 patients have been identified to date with mutations in *TAOK1* (Table 7), and they show a quite variable phenotype. Additionally, four more individuals with chromosomal deletions including *TAOK1* have been reported, but the genomic alterations affect other genes as well. To improve the definition of the clinical features associated with *TAOK1* mutations, it would be necessary to characterize more patients with identified mutations in this gene^{27,246,248,253,270–273}.

Currently, most of the disorder-associated variants identified in this gene are classified as *de novo* loss-of-function mutations, but inherited variants have also been found in some cases, as indicated in Table 7. Interestingly, variants have been linked to reduced and increased levels of expression of *TAOK1*, in both cases altering neural morphology and migration, which suggest the importance of a tight regulation of *TAOK1* during neurodevelopment^{246,248,256,271,272}.

1.8.3.3. Orthologous genes of *TAOK1* in zebrafish.

TAOK1 has 278 orthologous sequences in multiple organisms, including *D. rerio*. The orthologous gene in this model organism, known as *taok1* (ENSDARG00000069603), can be found in chromosome 15 (chr15:25588357-25629046) and it maintains an 83 % homology with the human sequence. According to public databases, *taok1* has 3 protein coding transcripts that generate two protein isoforms of 1014 and 1023 aa²⁰⁶. The levels of *taok1* remain low for most of the development, except for the blastula and larvae stages where *taok1* shows a peak of expression²⁴³.

There is also another sequence is *Danio rerio* identified as a potential orthologous of *TAOK1*, ENSDARG00000098304, localized in chromosome 21 (chr21:40528788-40579807). However, this protein coding gene is not identified as a high confidence gene due to its low conservation during evolution²⁰⁶.

1.9. ETHICAL AND LEGAL CONSIDERATIONS REGARDING THE EXPERIMENTAL PROCEDURES OF THIS THESIS.

The ethical considerations regarding this project revolve around to the use of hiPSCs and animal models for research purposes. This section is dedicated to expose the current legislative framework that must be followed by all research projects based on the use of the previously mentioned models.

The use of hiPSCs for research has undoubtedly less ethical concerns than the use of hESCs, which have already been explored in this introduction. Nevertheless, since hiPSCs are derived from adult somatic cells, it is important that their obtention and use respects the current legislation and the personal rights of the individuals. In this context, research projects conducted with hiPSCs must follow the indications of the Spanish Biomedical Research Law (*Ley 14/2007*), which include to be reviewed by any of the authorized Research Ethics Committees in the country before being approved by the *Comisión de Garantías para la Donación y Utilización de Células y Tejidos Humanos*. Moreover, since hiPSCs are derived from somatic cells, the donor (or legal representative) of those must read and signed an Informed Consent to allow the obtention and use of the generated hiPSCs^{86,87}.

Regarding the use of animals for research purposes must be strictly regulated with the objective of allowing scientific progress without jeopardising animal welfare. In this context, the European Union has a specific legislation framework to regulate the use of animals with scientific purposes, which was approved on 2010 (*2010/63/EU*) and entered into force on 2013. This law is based on the 3Rs principle, that promotes reducing, reusing, and refining the

use of animals in research projects, to improve their welfare, and protect their rights. Additionally, in Spain there is also a specific legislation concerning the use of animals for research (*Real Decreto 53/2013*, *Real Decreto 1386/2018*) and the training requirements needed for their handling (*Orden ECC/566/2015*)^{274–277}.

The present thesis has been conducted following all the previously stated regulations with the maximum respect for the rights of the hiPSCs-donors and the used animals.

Some content from this Introduction has been derived or modified from the publication by Pensado-López et al. (2020)¹¹, and the details are provided in Annexe VI.

JUSTIFICATION AND OBJECTIVES

2. JUSTIFICATION AND OBJECTIVES

2.1. JUSTIFICATION.

Neurodevelopmental disorders exhibit a broad spectrum of phenotypic and genotypic diversity. For this reason, unravelling the molecular underpinnings of these pathologies has historically presented a great challenge for the scientific community. This scenario changed with the development novel strategies, such as genome editing and reprogramming technologies, which have revolutionized our ability to functionally address human pathologies. The development of *in vitro* and *in vivo* models for human diseases, particularly those of neurological origin, represents a significant opportunity, and newfound hope, for patients who were previously considered beyond recovery.

The use of hiPSCs as *in vitro* models offers a revolutionary advantage for the study of neurological diseases and disorders. For instance, hiPSCs have the capacity to differentiate into numerous cell types, including neurons and glial cells, mirroring the complexity of the human CNS *in vitro*. Also, since hiPSCs can be obtained from adult somatic tissues, it is possible to preserve the genetic background of an individual, while addressing the specific impact of genomic variants potentially associated to the disorder of interest.

Despite their advantages, preclinical *in vitro* models still have limitations to overcome, such as the lack of global complexity and the absence of behavioural responses, both key for the study of neurodevelopmental alterations. To overcome these limitations, *Danio rerio* has been used as a model to study the impact of genomic variants potentially ligated with neurodevelopmental alterations. The past decades have consolidated *D. rerio* as a suitable animal model in neuroscience due to its genetic and physiological similarities to humans. The rapid development of the species, along with the transparency of the embryos, simplify the genetic manipulation and characterization of their functional consequences. Additionally, the small size and prolific reproduction of zebrafish facilitate large-scale screening of potential drug candidates, accelerating the identification of compounds with therapeutic potential.

In summary, these models not only facilitate a deeper understanding of specific neurodevelopmental disorders, but also open new avenues for the application of innovative therapeutical strategies, including gene therapy and high-throughput screening of small molecules capable of reversing or alleviating the patient's distinct symptoms. Consequently, this approach signifies a paradigm shift towards personalized and precision diagnosis and therapy—a concept that was scarcely imaginable until just a few years ago.

2.2. OBJECTIVES.

The main objective of the present doctoral thesis is to evaluate the applicability of functional assessment of potential new ASD risk genomic variants. In this thesis, three genes of interest were selected (*BTBD8*, *FOSL2*, and *TAOK1*) to evaluate their role human neurodevelopment and its alterations.

The specific objectives of the doctoral thesis are orientated to the generation of suitable models for each gene:

1. To generate and to characterize an *in vitro* hiPSCs-model to study *BTBD8*'s function.
 - 1.1. To develop a *BTBD8*-mutant hiPSCs line using the CRISPR-Cas9 system.
 - 1.2. To characterize the generated *BTBD8*-mutant hiPSCs line.
 - 1.3. To evaluate the impact of *BTBD8* mutations in human brain organoid development, focusing of major morphological alterations and cell-type composition.
 - 1.4. To assess the morphometric and proteomic impact of *BTBD8* mutations in the differentiation of human cortical neurons.
2. To generate and to characterize a knockdown zebrafish model to study the function of *fosl2*.
 - 2.1. To evaluate the viability impact linked to *fosl2*'s knockdown.
 - 2.2. To verify the knockdown in *fosl2* morphant zebrafish.
 - 2.3. To characterize the phenotype of *fosl2* morphant zebrafish.
3. To generate and to characterize *in vitro* and *in vivo* models to *TAOK1*'s function.
 - 3.1. To generate an hiPSCs model to study the function of *TAOK1*.
 - 3.1.1. To develop a *TAOK1*-mutant hiPSCs line using the CRISPR-Cas9 system.
 - 3.1.2. To characterize the generated *TAOK1*-mutant hiPSCs line.
 - 3.1.3. To evaluate the impact of *TAOK1* mutations in the human brain organoid development, combining information obtained from immunostaining and scRNA-seq.
 - 3.2. To generate a zebrafish model to study *taok1*'s function.
 - 3.2.1. To develop a zebrafish mutant line for *taok1* using CRISPR-Cas9 technology.
 - 3.2.2. To characterize the phenotype of the mutant embryos.

MATERIALS AND METHODS

3. MATERIALS AND METHODS

3.1. PROTOCOLS USED FOR *IN VITRO* EXPERIMENTS.

The procedures described in this doctoral thesis regarding the use of hiPSCs were conducted within the research project *Enfoque integral y translacional del diagnóstico y tratamiento de trastornos del neurodesarrollo (TEA, TDAH, DI) y del TOC: Una aproximación genética y funcional*. This project has been reviewed and approved by the Ethical Committee of Santiago-Lugo Research dependent of the *Consellería de Sanidade da Xunta de Galicia* (Ref. 2020/400, Annexe IV). In addition, the *Banco Nacional de Líneas Celulares, Instituto de Salud Carlos III* has reviewed and approved the use of the human induced pluripotent stem cell line Ctrl.BS-PBiPS37-Sv4F-1 in this research project (MTA displayed in Annexe V). All procedures were performed according to current legislation (Ley 14/2007).

3.1.1. Culture of human pluripotent stem cell lines.

Originally, hiPSCs cells were maintained *in vitro* on feeder-dependent conditions, using inactivated murine embryonic fibroblasts (MEFs) as a supporting layer. Nowadays, hiPSCs can also be maintained in feeder-free conditions, avoiding the need of using MEFs, due to the development of novel media formulations and matrices to cover culture plates. In this section the reader can find a summary of the basic maintenance conditions of the hiPSCs lines used in this thesis, which can be found in Table 8.

Table 8. hiPSCs lines used in this thesis. Data was obtained from the Spanish National Bank of Cell Lines²⁷⁸.

CELL NAME	LINE	PATHOLOGICAL CONDITIONS	CELL TYPE AND REPROGRAMMING METHOD	AGE	SEX	DEPOSITOR
Ctrl.BS-PBiPS37-Sv4F-1		None	PBMCs, Sendai virus	55	XY	Institut d'Investigació Biomèdica de Girona

3.1.1.1. Maintenance of hiPSCs.

In our lab, hiPSCs were maintained in feeder-free conditions with mTeSR Plus medium (StemCell Technologies) on plates coated with Matrigel (Corning). Matrigel coated plates were prepared by diluting Matrigel in DMEM/F-12 to a final concentration of 0.5 µg/cm². Plates were incubated for a minimum of 30 minutes (min) at 37 °C before using them.

Medium changes were performed daily and the health of the hiPSCs colonies was checked using a Leica DMI1 microscope (Leica Microsystems) to identify spontaneous differentiation or contamination in the cultures (Figure 17).

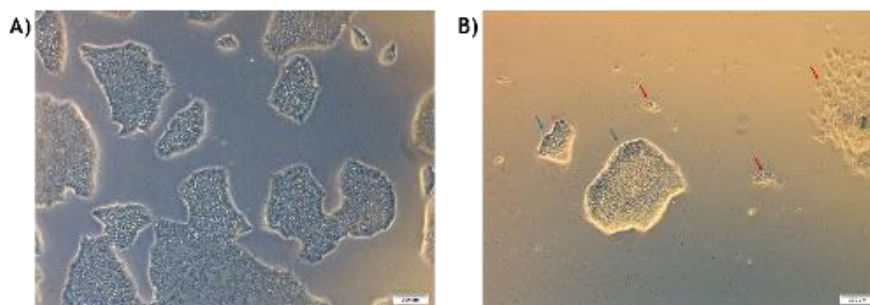


Figure 17. hiPSCs in culture. A) Expected morphology of hiPSC colonies. B) hiPSCs culture showing differentiation spots (red arrows) mixed with colonies with an appropriate shape (blue arrows).

3.1.1.2. Expansion of hiPSCs.

Before the expansion of any hiPSCs culture, it's important to check the health of the colonies under a microscope to confirm the absence of differentiation, an adequate confluency level (between 70 - 80 %), and the presence of compact colonies with defined edges.

In this thesis, two different methods were used to split hiPSCs cultures, depending on the need of seeding cells in a single cell form or not.

As a routine, hiPSCs were split using a non-enzymatic reagent called EDTA. This compound can disrupt hiPSCs' adhesion by chelating calcium, thus altering the function of cell adhesion proteins such as cadherins. This is a gentle dissociation process that results in the formation of small cell aggregates which have a better post-passage cell survival. The protocol for splitting cells with EDTA required the following steps. After coating cell culture plates with Matrigel (as described in Maintenance of hiPSCs.), cells were washed once with PBS without calcium and magnesium (-/-). Next, cells were incubated with EDTA (0.5 mM) for 4 min at 37 °C, after which EDTA was aspirated and cells resuspended in mTeSR Plus with Rock Inhibitor (Y-27632, RI) (Selleckchem) to a final concentration of 10 µM. Then, cells were counted in a Luna-II Automated Cell Counter (Logos Biosystems Inc) by mixing, in a 1:1 ratio, resuspended cells and Trypan Blue (0.4 %). Finally, the desired number of cells were seeded on Matrigel-coated wells in mTeSR Plus with RI (10 µM).

However, cells also needed to be seeded in a single cell form for some of the protocols described in this thesis, for example for differentiations and clone isolations. When this was the case, cells were split using Accutase, which relies on enzymatic dissociation. The protocol for passaging cells with Accutase also started with coating culture plates with Matrigel and washing cells with PBS -/-. Then, cells were incubated with Accutase for 5 min at 37 °C, after which a cell suspension was obtained. Accutase was neutralized with mTeSR Plus with RI (10 µM), and the solution was centrifuged for 4 min at 200 xg. Afterwards, the cell pellet was resuspended in fresh mTeSR Plus with RI (10 µM) and cells were counted and seeded as previously described.

3.1.1.3. Freezing and thawing hiPSCs.

Briefly, hiPSCs cultures were incubated with EDTA 0.5 mM for 5 min at 37°C. After that, the EDTA solution was aspirated, and the cells were recovered with mTeSR Plus with RI (10 µM). Cells were counted using an automated cell counter (Luna II), and for each cryovial to be frozen, an aliquot of 10^6 cells was taken and spun down for 3 min at 200 xg. After discarding the supernatant, the cell pellet was resuspended in 0.5 mL of KOSR with DMSO (10 %) and transferred to a cryovial that was put inside a Mr. Frosty (Thermo Fisher Scientific).

Cryovials were maintained for 24 hours (h) at least in a -80 °C freezer before transferring them into a liquid nitrogen tank for long-term storage.

Each cryovial, containing 1 million cells, was thawed in a water bath at 37 °C. Then, the content of the tube was mixed with 4 mL of mTeSR Plus with RI (10 µM), and centrifuged 5 min at 200 xg. The cell pellet was resuspended in 6 mL of mTeSR Plus with RI (10 µM), and then divided between three wells of a 6-well plate previously coated with Matrigel.

3.1.2. Gene knockout in hiPSCs.

Nowadays, there are the several options available for the delivery of CRISPR-Cas9 components into hiPSCs as previously mentioned in the introduction. In this thesis, a chemical mediated transfection of a plasmid containing the CRISPR-Cas9 components was selected since a detail protocol for knocking-out genes with this method was available²⁷⁹. Also, this method is non-integrative and transfected cells can be selected, which reduces the number of hiPSCs clones that need to be analysed before finding an edited one. A schematic summary of the gene knockout method that was used can be found in the Figure 18.

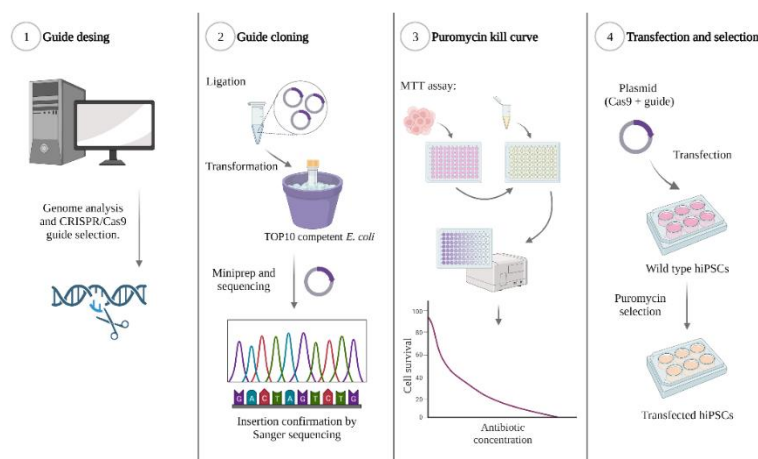


Figure 18. Gene knockout strategy followed in this thesis. Image created with Biorender.

3.1.2.1. Knockout guide design.

The design of the guides to knock-out our genes of interest in hiPSCs was performed using the webtools CRISPR-ERA v1.2²⁸⁰ and Synthego's Knockout Guide Design v1.3²⁸¹. Considering that the strategy for delivery of the CRISPR-Cas9 components was plasmid transfection, the guides with a predicted good expression under the U6 promoter were first identified. Then, they were classified attending to their proximity to the beginning of the gene, presence in all the protein-coding transcripts, predicted activity and minimal off-target effects. After applying these filters, the top three scoring guides were chosen per gene to be knocked out (*TAOK1* and *BTBD8*). The list of potential off-targets was obtained using the webtools Cas-OFFinder⁶⁸ and Synthego's Verify Guide Design²⁸². All the designs were performed using the GRCh38 genome assembly.

3.1.2.2. Cloning guides into plasmid vectors.

After choosing the guides to target our genes of interest in hiPSCs, the guide sequence was adapted to be cloned into a suitable plasmid, that includes the Cas9 coding sequence and a

region to clone the desired sgRNA, under the regulation of a suitable promoter for human cells. We chose to use the plasmid pSpCas9(BB)-2A-Puro (PX459) V2.0 (Figure 19), as it has been previously recommended by other authors²⁷⁹. This plasmid was a gift from Feng Zhang and was obtained through the Addgene repository (plasmid #62988)⁵⁰.

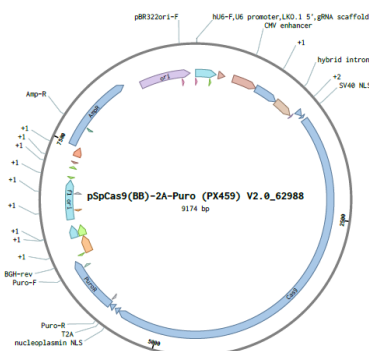


Figure 19. pSpCas9(BB)-2A-Puro (pX459) V2.0 schematic representation (Addgene #62988)⁵⁰. Image created by the author.

3.1.2.2.1. Plasmid purification.

Plasmids were obtained from Addgene repository were sent in *Escherichia coli* as stab agar, so it was necessary to expand and isolate them before being able to use them. First, bacteria were spread in LB agar plates supplemented with ampicillin (100 µg/mL) and the plates were left overnight at 37 °C. The next morning, bacteria were put in liquid culture to expand them and increase the amount of plasmid to be recovered. To do so, a single colony was picked and put in 5 mL of LB supplemented with ampicillin (100 µg/mL) and was left for 8 h in agitation at 37 °C. After that, the culture was mixed with other 250 mL of LB medium with ampicillin and incubated overnight at 37 °C with agitation. The following day, the culture was ready to be used for the purification of the plasmid of interest, for which the NucleoBond Xtra Midi kit for transfection-grade plasmid DNA (Macherey-Nagel) was used, following the steps described in the protocol.

3.1.2.2.2. sgRNA cloning.

To be able to clone the designed guides into the plasmid of interest (Figure 19), it is necessary to transform them into a double-strand DNA sequence that can be cloned into the plasmid. For this reason, compatible ends were included in the oligonucleotides, as indicated in Figure 20 and sequences were purchased from Integrated DNA Technologies (IDT).

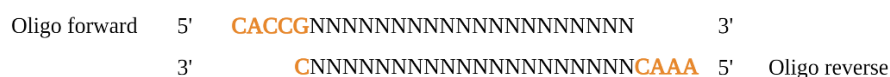


Figure 20. Oligonucleotides designed to be compatible with cloning into SpCas9 (BB)-2A-Puro (pX459) V2.0. In orange, compatible ends. N represent a generic 20 nucleotide sgRNA sequence.

Oligonucleotides were resuspended to 100 µM and annealed in the lab, by mixing 1 µL of each one with 1 µL of 10X T4 DNA Ligase Reaction Buffer, 0.5 µL of T4 Polynucleotide Kinase and 6.5 µL of nuclease-free water. Reactions were incubated for 30 min at 37 °C, followed by a 5-minute incubation at 95 °C and a final ramp down to 25 °C (decreasing 5

°C/min). For each guide, the generated oligo duplex was diluted 1:250 and then ligated to the plasmid pSpCas9(BB)-2A-Puro (pX459) V2.0. The plasmid digestion and the ligation of the oligo duplex for each guide was performed in just one step, by mixing the reagents indicated in Table 9, as previously described²⁷⁹.

Table 9. One-step digestion and ligation of guides into plasmid.

REAGENT	STOCK CONCENTRATION	VOLUME PER REACTION (μL)
Plasmid (Addgene #62988)	100 ng/μL	1 μL (100 ng)
Guide oligo duplex	1:250 dilution	2 μL
Fast Digest Buffer	10X	2 μL
DTT	10 mM	1 μL
ATP	10 mM	1 μL
Fast Digest <i>Bpi</i> I		1 μL
T7 DNA ligase	3000 units/μL (U/μL)	0.5 μL
Nuclease-free water	-	11.5 μL

Reactions were incubated for 5 min at 37 °C, followed by an incubation of 5 min at 23 °C. This cycle was repeated 6 times, and then, the reaction was treated with Plasmid-Safe exonuclease to eliminate linear DNA from the reaction. To do so, 11 μL of reaction were mixed with 1.5 μL of Plasmid-Safe Buffer, 1.5 μL of ATP and 1 μL of Plasmid-Safe DNase. The reaction was incubated 30 min at 37 °C, after which it was ready to be used in the transformation of competent bacteria, which is a necessary step in to identify plasmids that have been cloned without errors.

3.1.2.2.3. Transformation, purification and plasmid sequencing.

For each guide, the ligation reaction obtained in the previous subheading is a mixture of recombinant plasmids, some of which contain the guide correctly cloned. To identify those plasmids, it is necessary to isolate and sequence them, to verify the absence of errors in the guide sequence which would alter the results of the CRISPR-Cas9 experiment.

To be able to do that, first of all, each ligation reaction was transformed into competent *E. coli*, following the steps described in Figure 21.

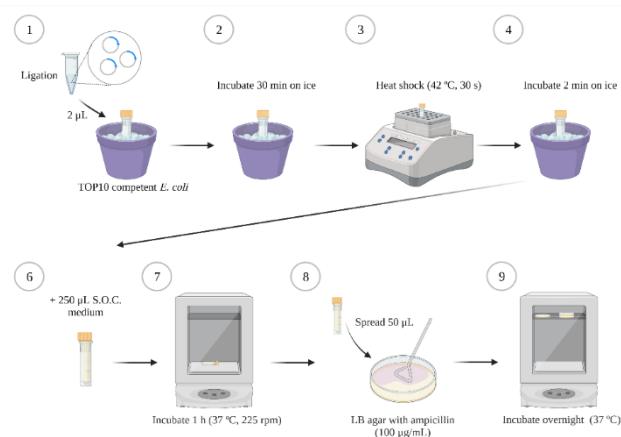


Figure 21. Steps require to transform plasmids into *E. coli*. Image created with Biorender.

The day after plating the bacteria into LB agar plates with ampicillin (100 μg/mL), 6 colonies were picked for each transformation and cultured in 3 mL of liquid LB with ampicillin (100 μg/mL), which were grown overnight at 37 °C in agitation. Next morning, plasmids were

purified from bacteria cultures using the NucleoSpin Plasmid Mini Kit (Macherey-Nagel) following the protocol of the kit.

After the purification, plasmids were diluted to 100 ng/ μ L and sequenced using the hU6 sequencing primer (5'- ACTATCATATGCTTACCGTAAC -3') at 10 μ M. Sanger sequencing was performed in the *Fundación Pública Galega de Medicina Xenómica* (Santiago de Compostela) and sequencing results were analysed using the software CodonCode Aligner v10.0.2 (CodonCode Corporation).

3.1.2.3. Puromycin kill curve for the hiPSCs line Ctrl.BS-PBiPS37-Sv4F-1.

As previously mentioned, the selected approach for knockout generation in this thesis consists in the transformation of a plasmid that contains the CRISPR-Cas9 components, as well as a puromycin resistance gene (Figure 19) that will give resistance to antibiotic treatment to the transfected cells. For this reason, it is important to first determine the basal sensitivity to puromycin treatment of our wild type (WT) hiPSCs line (Ctrl.BS-PBiPS37-Sv4F-1).

The puromycin kill curve for the Ctrl.BS-PBiPS37-Sv4F-1 cells was performed for a total of 5 days using antibiotic concentrations ranging between 0 and 1.5 μ g/mL, to establish the dose-effect response of the cell line to puromycin.

Briefly, 96-well plates were coated with Matrigel as described beforehand (see Maintenance of hiPSCs) and 2500 cells were seeded per well in 100 μ L of mTeSR Plus with RI (10 μ M) for 24 h. Three replicas were done per concentration, including a blank condition covered with Matrigel and filled with medium but without cells. The next day, medium was changed with mTeSR Plus with puromycin at the selected concentrations, step that was repeated daily until the end of the experiment. Finally, cell viability was measured by a MTT assay.

To perform the viability assay, 10 μ L of MTT (5 mg/mL in PBS) were dispensed to each 96-well and incubated 4 h at 37 °C. Then, 100 μ L of 10 % SDS (in HCl 0.01 M) were added per well and incubated again for 14 h at 37 °C. After that, the absorbance was measured in a Mithras LB 940 (Berthold Technologies) plate reader at 590 nm. Data analysis was performed in GraphPad Prism 8.0.2 and survival curves were obtained by applying a nonlinear regression analysis, as described by Delrue *et al.*²⁸³.

3.1.2.4. Transfection of CRISPR-Cas9 components and antibiotic selection.

The day before transfection, 450000 WT cells (hiPSC line Ctrl.BS-PBiPS37-Sv4F-1) were seeded per each well of a 6-well plate previously coated with Matrigel in mTeSR Plus with RI (10 μ M). After 24 h, medium was changed again with mTeSR Plus with RI (10 μ M), and the transfection complexes were prepared. To do so, in one tube, 6 μ L of Lipofectamine Stem Transfection Reagent (Invitrogen) were mixed with 100 μ L of Opti-MEM (Gibco). In another tube, 2 μ g of plasmid were added to 100 μ L of Opti-MEM. Then, this second solution was mixed with the first one and incubated for 10 min at room temperature. Finally, the complexes were added dropwise to the cells, and plates were incubated overnight at 37 °C.

For the next two days, medium was changed with mTeSR Plus with RI (10 μ M) and puromycin to kill the non-transfected cells. After the selection, cells were fed again with mTeSR Plus with RI (10 μ M) for two more days.

Finally, medium was changed with just mTeSR Plus until the cells were confluent enough to be split. At that point, cells were dissociated with Accutase, and half of them were frozen, while the other half was used to analyse the guide efficiency as described in the following subheading.

3.1.3. Guide efficiency analysis.

The fact that transfected cells were selected doesn't mean that the surviving cells were edited. This, added to the fact that several guides were tested for each knockout experiment, makes it relevant to analyse how efficient the editing process was. This information will be relevant to continue the experimental workflow with the most efficient guide, to maximize the chances of obtaining full knockout hiPSCs clones.

In this thesis, two approaches were used to analyse the efficiency of the designed guides: Interference of CRISPR Edits (ICE) analysis and TOPO cloning, summarized in Figure 22.

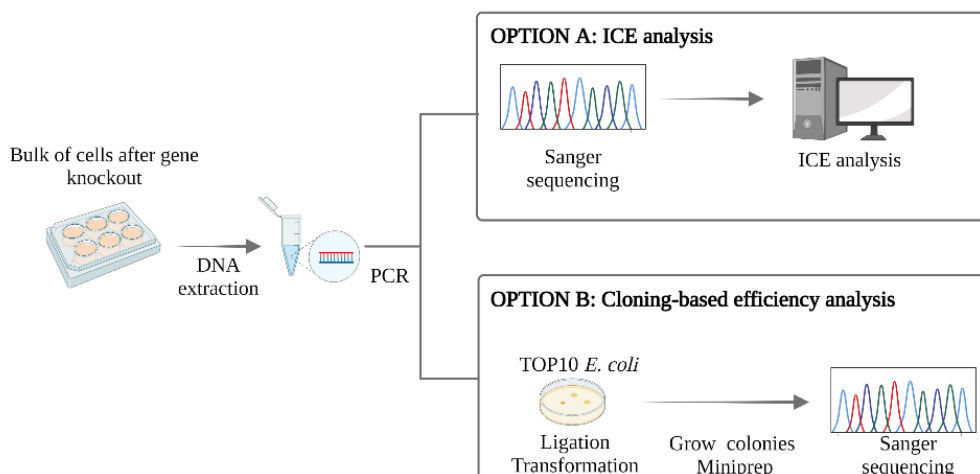


Figure 22. Guide efficiency analysis methods used in this thesis. Image created with Biorender.

3.1.3.1. DNA extraction from hiPSCs.

The first step to check if the CRISPR-Cas9 experiment worked consists in the extraction of DNA from part of the bulk of transfected and selected cells. To do so, cells were pelleted by centrifugation for 5 min at 300 xg and then the DNA extractions were performed using the DNeasy Blood and Tissue Kit (Qiagen), following the instructions of the kit.

3.1.3.2. Amplification of the target region for CRISPR-Cas9 in the bulk of transfected cells.

The DNA that was extracted in the previous subheading was used as a template for a PCR reaction to amplify the target region for each guide to detect the presence of mutations. To avoid detecting mistakes introduced by the polymerase, a high-fidelity enzyme was used. Each PCR was prepared by mixing 12.65 μL of nuclease-free water, 4 μL of High-Fidelity Buffer, 0.4 μL of 10 mM dNTPs, 1 μL of each primer at 10 μM , 0.2 μL of Phusion High-Fidelity DNA Polymerase, and 1 μL of DNA template at 50 ng/ μL . The general amplification conditions can be found in the Figure 23. The primer pairs used for sequencing the regions of interest can be found in the Table 10, as well as the specific amplification conditions needed for each pair.

Table 10. Sequencing primers for the identification of genetic edition in hiPSCs.

GENE	TARGET EXON		PRIMER SEQUENCE (5'-3')	AMPLICON SIZE (bp)	ANNEALING (°C)	EXTENSION TIME (s)
TAOK1	2	Forward	ACCCACACTGCCTTTTCAGT	1027	65 °C	35
		Reverse	CTCTATAGCAATGCAAGCCATGAGA			
TAOK1	4	Forward	ATGGTGGTGTGCGCCTGTAGAC	570	67 °C	30
		Reverse	TCTGCAGTGAGACTTGCAAAAGTTC			
BTBD8	5	Forward	GGAGCATCTGGACATTCTGTC	892	60 °C	30
		Reverse	CTCTTGAAACAGGTCGAGAT			
BTBD8	7	Forward	CTATGGCATATGATTATACCTTGCT	523	62 °C	30
		Reverse	CCCAAAACACCTGTCTGAAAAG			

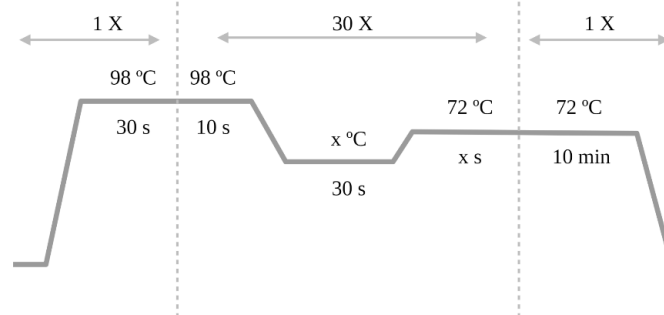


Figure 23. Program used for PCR with hiPSCs samples using Phusion polymerase.

Amplification was detected by gel electrophoresis running the samples at 100 V for 40 min in 1 % agarose gels, prepared in TAE buffer with GelRed (Biotium). A 100 bp DNA ladder was included in the gels and samples were loaded by mixing them with 6 X loading buffer.

3.1.3.3. ICE analysis.

The PCRs obtained from the different bulks of transfected cells were sequenced using the Sanger method in the *Fundación Pública Galega de Medicina Xenómica*, using the same primers that were used in the amplification step (see Table 10). Sequencing data was then analysed by the ICE analysis tool offered by Synthego²⁸⁴, which uses Sanger sequencing data to estimate the editing efficiency in a sample.

3.1.3.4. Cloning-based efficiency analysis.

The other approach that was used in this thesis to determine the CRISPR-Cas9 efficiency was also based on sequencing the PCR fragments from the subheading 3.1.3.2. However, in this case, they were first isolated and amplified in bacteria to get a clear read-out of the types of mutations generated by the editing system.

Previously to performing the cloning step, three replicas of each PCR described on the subheading 3.1.3.2 were ran together in the same well of an agarose gel. After the electrophoresis, each amplification band was extracted by cutting the agarose and then purifying it using the QIAquick Gel Extraction Kit (QIAGEN), following the indications of the company.

After purification, the Zero Blunt TOPO PCR Cloning kit (ThermoFisher) was used to ligate 4 µL of PCR product with 1 µL of vector and 1 µL of salt solution. The ligation reaction was mixed and incubated for 5 min at room temperature before transforming it into competent bacteria. For each transformation, 2 µL of ligation were mixed with one vial of TOP10 chemically competent *E. coli* and incubated on ice for 30 min, after which a heat shock (30

seconds, 42 °C) was applied. Then, 250 µL of S.O.C medium was added to the bacteria, and they were incubated for 1 hour at 37 °C with agitation (230 rpm). Next, 30 µL of bacteria were seeded on a LB agar plate supplemented with kanamycin (50 µg/mL) and incubate overnight at 37 °C.

The next day, six colonies were picked per transformation and cultured in 3 mL of LB with kanamycin (50 µg/mL). Liquid cultures were grown overnight at 37 °C with agitation (230 rpm).

The following day, the transformed plasmids were isolated from the liquid cultures using the QIAquick Spin Miniprep kit (Qiagen). Each purified plasmid was quantified in a Nanodrop and prepared to be sequenced.

The M13 primer (5'-TGTAACGACGGCCAGT-3') was used for sequencing since it binds near the cloning region and the insert is the only part of the plasmid that we wanted to analyse. Samples were prepared to be sequenced by mixing 5 µL of M13 primer (5 µM) with 5 µL of plasmid (75 ng/µL). Sanger sequencing was conducted using the service Mix2Seq by Eurofins Genomics. Sequencing data was analysed using the software SnapGene and CodonCode Aligner v10.0.2.

3.1.4. Isolation and characterization of edited clones.

As previously mentioned, the bulk of cells obtained after a gene knockout experiment performed by CRISPR-Cas9 is a mixture of cells with different types of edited and non-edited alleles. So, to obtain a knockout cell line, it's important to first isolate homogeneous clones grown from just one cell, and then analyse if they have the mutations of interest and remain pluripotent.

Finally, modified cell lines were named following the guidelines of the *Banco Nacional de Líneas Celulares and the Instituto de Salud Carlos III*²⁸⁵. Those guidelines indicate that the name of a cell line derived from a previously banked hiPSCs line should include: the name of the original hiPSCs line, followed by the modified gene, and an index that indicates if it has one or two modified alleles (i1 for one allele, i2 for two alleles).

3.1.4.1. Isolation of edited clones.

To be able to isolate edited clones, it's necessary to plate hiPSCs at low density. This will allow single cells to form colonies that will be genetically homogeneous.

To do so, hiPSCs were detached with Accutase as previously indicated, and 50 cells were seeded in 6-well plates coated with rhLaminin521 at 10 µg/mL. To improve cell survival, cells were seeded in mTeSR Plus supplemented with 1/10 CloneR2 (StemCell Technologies), which was be refreshed two days after. After two more days, regular medium changes were performed until colonies were big enough to be picked (see Figure 24).

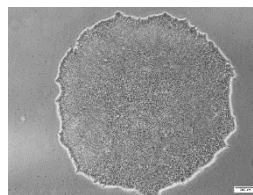


Figure 24. hiPSC colony grown from a single cell.

At that time point, colonies were marked and then picked using a 100 μ L tip under a loupe. Each colony was seeded in a well of a 24-well plate coated with Matrigel in mTeSR Plus medium supplemented with 1/10 CloneR2 (StemCell Technologies) and 1/100 Penicillin/Streptomycin (10000 U/mL) to avoid contamination. Two days after, medium was changed with mTeSR Plus supplemented with 1/100 Penicillin/Streptomycin (10000 U/mL), process that was repeated for the next three days. After that, regular medium changes were performed, as previously described. When the clones were confluent enough to be passaged, each one was seeded into a new well of a 12-well plate coated, following the usual protocol for hiPSCs maintenance (described in the subheading 3.1.1.1). When the clones reached 80 % confluency, half of the cells was frozen, and the other half was used to perform a DNA extraction to genetically characterize them.

3.1.4.2. Allele isolation and identification.

The first thing that was analysed on the isolated clones was the presence of mutations derived from the activity of the CRISPR-Cas9 system.

To do so, DNA was extracted as described previously in the subheading 3.1.3.1. Then, samples were amplified by PCR as indicated in 3.1.3.2 and the PCR fragments were cloned into bacteria to isolate each fragment on a different colony. This was conducted following the steps that are indicated in 3.1.3.4, however, in this case 20 colonies were analysed for each transformation.

3.1.4.3. Karyotyping of hiPSCs.

As hiPSCs lines show a propensity towards genomic instability during long-term culture, it's also important to check that edited clones have no karyotype differences comparing to their isogenic control.

To do so, a full 6-well plate was seeded per line to be karyotyped and cells were grown until they reached 80 % confluency. Then, to induce cell immobilization during metaphase, each well was incubated for 45 min at 37 °C with mTeSR Plus supplemented with 10 μ g/mL of N-desacetyl-N-methylcolchicine (Colcemid, ThermoFisher). After the incubation, cells were washed twice with PBS (-/-) and dissociated with Accutase to obtain a single cell suspension, as previously described. After washing the pellet once with PBS (-/-), cells were resuspended in 10 mL of hypotonic solution (KCl 0.075 M) while vortexing. This cell suspension was incubated for 10 min at 37 °C. While vortexing, 1 mL of cold Carnoy fixative (3:1, Methanol: Acetic Acid) was added to the suspension and then centrifugated at 562 xg for 10 min at room temperature. The supernatant was discarded, and the pellet resuspended while vortexing in 10 mL of cold Carnoy fixative. Finally, samples were sent to the Cytogenetics Unit of the *Fundación Pública Galega de Medicina Xenómica* for karyotype analysis, which performed by Luz Míguez Álvarez, PhD.

3.1.4.4. Pluripotency analysis by immunocytochemistry.

With the objective of analysing whether the generated edited hiPSCs lines remained pluripotent, immunocytochemistry (ICC) was used to detect the expression of the pluripotency markers NANOG, SOX2 and TRA-1-81.

First, cells were grown on SlideFlasks (Thermo Fisher Scientific) to form separated colonies. Then, after washing cells three times with PBS, they were fixed with PF4 4 %. After

washing the cells again, they were incubated for 1 hour in blocking solution, which is a solution of 5 % normal donkey serum (NDS) in 0.25 % triton-X-100 in PBS (TPBS). Once the blocking step was over, cells were incubated overnight at 4 °C in blocking solution with the primary antibodies for the pluripotency markers of interest (NANOG, TRA-1-81, SOX2, as indicated in Table 11).

The next morning, cells were washed twice with 0.25 % TPBS and then incubated for 5 min in blocking solution. After that step, cells were incubated for 2 h at room temperature in blocking solution with the secondary antibodies (Table 11) and Hoechst 33342 (Thermo Scientific, diluted 1:500) to stain the nuclei of the cells. Finally, after washing twice with PBS, the slides were submerged in milli-Q H₂O and coverslips were mounted with polyvinyl alcohol mounting medium with DABCO.

Images were taken with an Olympus BX61 Fluorescence Microscope and the software cellSens Dimension (Olympus Corporation), and then processed using ImageJ.

Table 11. Antibodies used for the pluripotency analysis of the hiPSCs cell lines.

ANTIBODY NAME	TYPE	REFERENCE	DILUTION
Human/Mouse/Rat SOX2 Antibody	Primary	AF2018 (R&D Systems)	1:100
Anti-Nanog antibody	Primary	ab21624 (abcam)	1:100
Anti-TRA-1-81 antibody	Primary	MAB4381 (Sigma-Aldrich)	1:200
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647	Secondary	A31573 (Invitrogen)	1:500
Alexa Fluor 488 AffiniPure Donkey Anti-Mouse IgM, μ chain specific	Secondary	715-545-140 (Jackson ImmunoResearch)	1:500
Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 647	Secondary	A21447 (Invitrogen)	1:500

3.1.4.5. Pluripotency analysis by RT-qPCR.

The pluripotency of the generated edited hiPSCs lines was also verified by real time quantitative PCR (RT-qPCR), analysing the expression the hiPSCs markers *NANOG*, *SOX2* and *POU5F1*.

3.1.4.5.1. RNA extraction from hiPSCs.

RNA was extracted from hiPSCs cultures using the RNeasy Mini kit (QIAGEN), following the instructions of the commercial handbook. Four biological replicas were obtained for each cell line, using hiPSCs cultures of different passage. RNA concentration was measured with a Nanodrop instrument.

3.1.4.5.2. cDNA synthesis.

To reverse-transcribe RNA into cDNA, the iScript cDNA Synthesis Kit (Bio-Rad) was used. Following the commercial instructions, reactions were prepared as indicated on Table 12. Briefly, reactions were incubated for 5 min at 25 °C in an annealing step, followed by a reverse-transcription step of 20 min at 46 °C, and a last step of 1 minute at 96 °C to inactivate the enzyme. Finally, samples were stored at -20 °C.

Table 12. cDNA synthesis reaction set-up.

REAGENT	VOLUME PER REACTION
iScript Reaction Mix 5X	4 μ l
iScript Reverse Transcriptase (RT)	1 μ l
RNA	0.9 μ g
Nuclease free water	Complete to 20 μ l

3.1.4.5.3. Desing and efficiency analysis of RT-qPCR primers.

Primers for RT-qPCR with SYBR Green chemistry were designed with Primer Blast and Primer3. In Table 13 can be found the sequences and amplicon size of the designed primer pairs.

The efficiency of the designed primer pairs was assessed by running a standard curve using 1/10 serial dilutions of WT cDNA as templates. The RT-qPCR reactions were prepared with the iTaq Universal SYBR Green Supermix kit as indicated in Table 14, with two technical replicates per condition. Reactions were run following the program in Figure 25 in a QuantStudio 5 (Applied Biosystems) instrument.

Table 13. Sequences and amplicon size of the RT-qPCR primers selected for pluripotency analysis.

GENE	FORWARD SEQUENCE (5->3)	REVERSE SEQUENCE (5->3)	AMPLICON SIZE
<i>GAPDH</i>	GAAGGTCGGAGTCAACGGATT	TGACGGTGCCATGGAATTTG	170 bp
<i>NANOG</i>	AGGAAGACAAGGTCCCGGTCAA	TCTGGAACCAGGTCTTCACCTGT	174 bp
<i>POU5F1</i>	GGGTTTTTGGGATTAAGTTCTTCA	GCCCCACCCTTTGTGTT	63 bp
<i>SOX2</i>	CAAAAATGGCCATGCAGGTT	AGTTGGGATCGAACAAAAGCTATT	63 bp

Table 14. RT-qPCR reaction set-up with SYBR Green chemistry.

REAGENT	VOLUME PER REACTION
iTaq Universal SYBR Green Supermix 2X	5 µl
Primer forward (5 µM)	0.8 µl
Primer reverse (5 µM)	0.8 µl
Nuclease-free water	2.4 µl
cDNA	1 µl

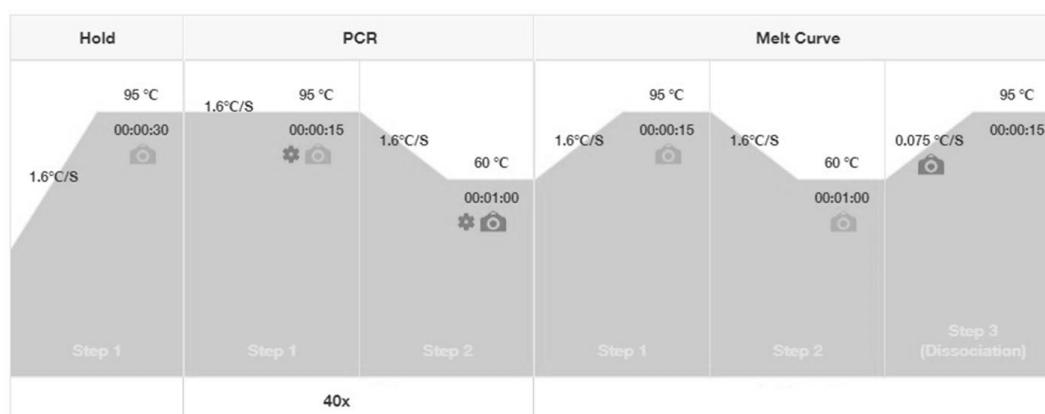


Figure 25. Program used for RT-qPCR with SYBR Green chemistry in the Quant Studio 5 (Applied Biosystems).

Finally, GraphPad Prism 8.0.2 was used to analyse the results. The linear regression line that adjusts to the data was calculated, and its corresponding equation obtained (see Equation 1), along with the goodness of fit (R^2). The slope of the linear regression line (indicated as “a”) on Equation 1, was used to calculate the amplification efficiency of each pair of primers, following the Equation 2. Reproducible data can be ensured when the efficiency falls within the range of 90 % to 110 %, otherwise it would be necessary to redesign the primers.

Equation 1. Example of the equation of a linear regression line. The slope line is represented as “a”.

$$y = ax + b$$

Equation 2. Formula used to calculate the amplification efficiency of the RT-qPCR primers.

$$\text{Efficiency (\%)} = (10^{(-1/a)} - 1) \times 100$$

3.1.4.5.4. RT-qPCR based on SYBR Green chemistry.

Once the primer pairs efficiency had been validated, they were used to perform RT-qPCRs. The reaction set-up and the amplification program were previously described (see Table 14 and Figure 25). For RT-qPCR experiments, at least three biological replicates were used for each experimental condition, always using three technical replicates for each sample.

Data was analysed in GraphPad Prism 8.0.2 following the delta-delta Ct method (ddCt) for relative expression quantification. Also, the normality of the data was analysed applying the D'Agostino & Pearson test. Differences between normally distributed groups were assessed with a two-tailed t-test. For non-normally distributed groups, the equivalent non-parametric test was applied (Welch's t-test). For every statistical analysis, the considered p-value was 0.05.

3.1.4.6. Off-target analysis in edited hiPSCs lines.

The guides used for CRISPR-Cas9 can also bind with several mismatches to genomic regions (off-targets) besides the target one. For this reason, when conducting a CRISPR-Cas9 experiment is important to validate that no mutations were incorporated in these off-target regions, since they could interfere with the interpretation of the experiments conducted with the generated cell lines.

3.1.4.6.1. DNA extraction.

DNA was extracted from hiPSCs cultures using the trizol/chloroform method. After removing the culture medium, between 10^6 and 10^7 cells were incubated for 5 min with 400 μ L of trizol (Thermo Fisher Scientific). Then, 80 μ L of chloroform was added to the sample and mixed using a vortex for 40 seconds.

After a 2-minute incubation, samples were centrifuged (15 min, 12000 xg, 4 °C). The obtained aqueous phase (containing the RNA) was discarded, and 120 μ L of 100 % ethanol was added to the remaining solution. Samples were mixed and incubated for a couple of min, and then centrifuged again (5 min, 2000 xg, 4 °C). As a result of this step, a pellet containing the DNA was formed, which was resuspended in 400 μ L of 0.1 M sodium citrate. Samples were incubated for 30 min, after which they were centrifuged once more (5 min, 2000 xg, 4 °C). This washing step was repeated twice, before resuspending the pellet in 600 μ L of 75 % ethanol.

After 15 min, samples were centrifuged (5 min, 2000 xg, 4°C), and the resulting pellet left for air drying around 10 min. The pellet was then resuspended in 300 μ L of 8 mM NaOH and centrifuged (10 min, 12000 xg, 4 °C). Finally, the supernatant was transferred to a new tube, pH was adjusted to 7.5 with HEPES, and EDTA 1mM was added to the DNA solution.

DNA concentration was measured using a Nanodrop instrument and samples were stored at -20 °C.

3.1.4.6.2. PCR and Sanger sequencing.

DNA from edited hiPSCs lines was used as a template for a PCR reaction to amplify the potential off-target regions after designing adequate sequencing primer pairs for

the regions of interest. Each PCR was prepared with 16.8 μL of nuclease-free water, 2.5 μL of Buffer II 10X, 1.5 μL of 25 mM MgCl_2 , 0.5 μL of 10 mM dNTPs, 0.5 μL of each primer at 10 μM , 0.2 μL of AmpliTaq DNA polymerase 5 U/ μL , and 2.5 μL of DNA template at 10 ng/ μL . The amplification conditions can be found in the Figure 23.

Samples were run in 1 % agarose gels, prepared in TBE buffer with RedSafe, and loaded with 6 X loading buffer. A 1 kb DNA ladder was also included in the gels, and electrophoresis was conducted for 40 min at 100 V.

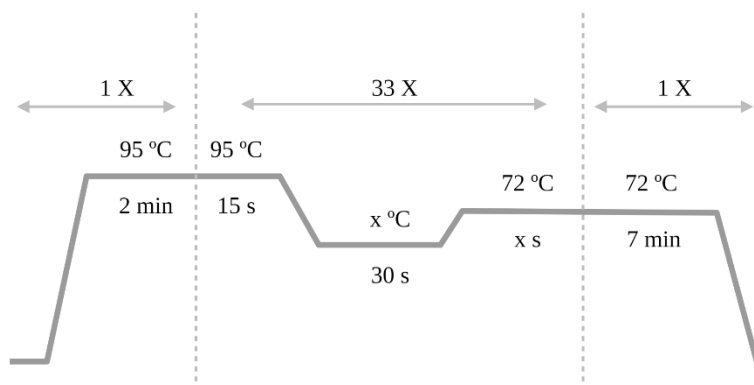


Figure 26. Program used for PCR with hiPSCs samples using AmpliTaq DNA polymerase. Image created using BioRender.com.

Samples were sequenced using the Sanger method in the *Fundación Pública Galega de Medicina Xenómica*. Sequencing data was analysed using the software CodonCode Aligner v10.0.2 (CodonCode Corporation).

3.1.4.7. Trilineage differentiation.

The ability of the edited hiPSCs to differentiate to the three germ layers was validated with the STEMdiff Trilineage Differentiation kit (StemCell Technologies).

Following the indications of the commercial handbook, hiPSCs were dissociated to single cells with Accutase as previously described (see 3.1.1.2). Afterwards, cells were seeded on 24-well plates, adding 100000 cells/well for those to be differentiated towards mesoderm, and 400000 cells/well for both ectoderm and endoderm.

On the day after seeding (day 1), mTeSR plus medium was replaced with the appropriate differentiation medium for each germ layer. This step was repeated daily until day 5 for mesoderm and endoderm lineages, while for ectoderm lineages it was maintained until day 7.

The final day of the protocol, cells were harvested, and RNA was extracted for the verification of the expression of specific markers for each type of lineage by RT-qPCR. The protocols required for it were performed as indicated in section 3.1.4.5, while the markers selected for this purpose are indicated in Table 15.

Table 15. RT-qPCR primer pairs designed for the amplification of specific markers for each germ layer.

GENE	FORWARD SEQUENCE (5->3)	REVERSE SEQUENCE (5->3)	AMPLICON SIZE	GERM LAYER
GAPDH	GAAGGTCGGAGTCAACGGATT	TGACGGTGCCATGGAATTTG	170 bp	Housekeeping
AFP	ACTCCAGTAAACCCTGGTGTTG	GAAATCTGCAATGACAGCCTCA	255 bp	Endoderm
NCAM1	TGTCCGATTCATAGTCCTGTCC	CTCACAGCGATAAGTGCCCTC	82 bp	Mesoderm
NES	CAGCGTTGGAACAGAGGTTGG	TGGCACAGGTGTCTCAAGGGTAG	389 bp	Ectoderm

3.1.5. Differentiation of human pluripotent stem cell lines.

3.1.5.1. Differentiation of hiPSCs to cortical-like neurons.

The obtention of cortical-like neurons from hiPSCs was conducted following the protocol published by Shi *et al.*¹⁰⁵, which is represented in Figure 27, with some minor modifications. This protocol uses dual SMAD inhibition to drive cells towards neuronal differentiation. The SMAD inhibitors used in this protocol are SB431542 and Noggin. This first one is an inhibitor of ACVR1B, TGFBR1, and ACVR1C, thus inhibiting the TGF- β /Activin/NODAL pathway, while Noggin acts as a BMP inhibitor. Dual SMAD inhibition is a well-known method to obtain neural progenitor cells in monolayers, which can later be matured into deep- and upper- layer excitatory neurons.

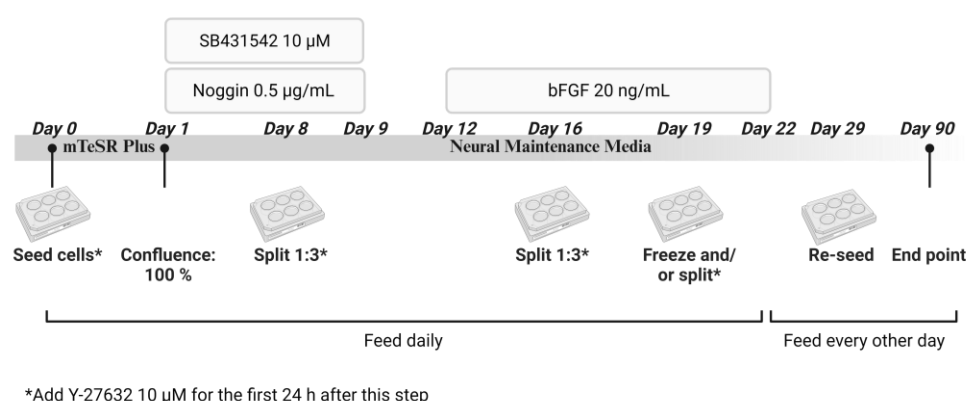


Figure 27. Schematic representation of the differentiation protocol used for the generation of cortical-like neurons, derived from Shi *et al.*¹⁰⁵ Image created with Biorender.com.

Table 16. Components of the neural maintenance medium (NMM) used for the differentiation of cortical-like neurons.

REAGENT	REFERENCE	STOCK CONCENTRATION	VOLUME TO PREPARE 500 mL
DMEM/F-12, GlutaMAX supplement	10565018 (Gibco)	100 %	250 mL
Neurobasal medium	21103049 (Gibco)	100 %	250 mL
N-2 Supplement	17502048 (Gibco)	100 X	2.5 mL
B-27 Supplement, serum free	17504044 (Gibco)	50 X	5 mL
GlutaMAX Supplement	35050061 (Gibco)	100 X	2.5 mL
Insulin-Transferrin-Selenium-Sodium Pyruvate (ITS-A)	51300044 (Gibco)	100 X	2.5 mL
Penicillin-Streptomycin	15140122 (Gibco)	10000 U/mL	5 mL
2-mercaptoethanol	21985023 (Gibco)	55 mM	400 μ L
MEM Non-Essential amino acids solution	11140050 (Gibco)	100 X	2.5 mL

The first part of the protocol consisted in the generation of neural stem cells (NSCs) from hiPSCs. On day 0, 1-well of a 6-well plate with 80 % confluent hiPSCs was dissociated with EDTA 0.5 mM (as described previously on 3.1.1.2) and the cells were seeded in a new well of

a 6-well plate in mTeSR plus medium supplemented with 10 μ M Y-27632. On day 1 the medium was substituted with neural maintenance medium (NMM, see Table 16) supplemented with 10 μ M SB431542 (Tocris Biosciences) and 0.5 μ g/mL Noggin (R&D Systems). This media change was repeated daily until day 8. During this time, cells should form a tight uniform neuroepithelial sheet. On day 8, cells were dissociated with EDTA 0.5 mM (4 min at 37 °C) and then resuspended in Neural Maintenance Medium (NMM) supplemented with 10 μ M SB431542, 0.5 μ g/mL Noggin, and 10 μ M Y-27632. The content was split 1 to 3 into Matrigel-coated wells of a 6-well plate.

Starting on day 9, daily medium changes were performed just with NMM. This continued until day 12, when the NMM began to be supplemented with 20 ng/ μ L bFGF. On day 16, cells were dissociated with Accutase, as described in 3.1.1.2, and split once again in a 1:3 ratio in Matrigel-coated wells of a 6-well plate. On this day, cells were seeded in NMM supplemented with 20 ng/ μ L bFGF and 10 μ M Y-27632 for 24 h.

On day 19, NSCs (Figure 28) were harvested to make frozen stocks. To do so, the cell culture was dissociated with Accutase (see 3.1.1.2) and 2 million cells were placed on each cryovial in freezing medium (NMM supplemented with 10 % DMSO).

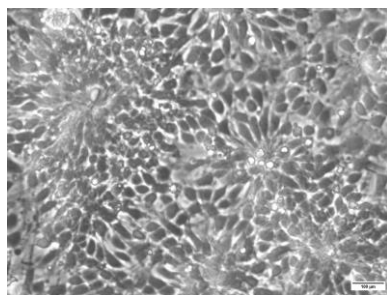


Figure 28. Neural rosettes formed *in vitro* after 19 days of cortical differentiation.

The second part of the protocol was the differentiation of NSCs to mature neurons. This step began after thawing NSCs in NMM supplemented with 20 ng/ μ L bFGF and 10 μ M Y-27632 in Matrigel-coated 6-well plates, at a seeding density of 50000 cells/cm². Medium was changed daily with NMM supplemented with 20 ng/ μ L bFGF. When 100 % confluence was reached (typically two days after seeding), medium began to be changed every other day with NMM only until the end of the protocol.

It is also important to mention that between day 28 and 30, cells were split a final time with Accutase in a 1 to 4 ratio. Although this step was optional, cell viability and long-term survival of the generated neurons was improved by its application.

3.1.5.2. Differentiation of hiPSCs to human brain organoids.

In order to generate human dorsal forebrain organoids, in this thesis we followed the protocol described by Velasco *et al.*^{119,286}, which is summarized in Figure 29.

Briefly, hiPSCs lines (all below passage 50) were cultured to reach 80-90 % confluency on day 0, at which point they were dissociated with Accutase, as described beforehand. The resulting cell pellet was resuspended in 1 mL of seeding medium and cells were centrifuged (5 min at 300 xg), step that was repeated twice. After counting, 9000 cells were added on each well of a Prime surface 96V plate (Sbio Japan) in 100 μ L of seeding medium. On this day, cells were seeded in Cortical Differentiation Medium 1 (CDMI, which recipe can be found on Table 17), supplemented with RI 20 μ M, IWR1 3 μ M and SB431542 5 μ M. Finally, the plate was centrifuged at 200 xg for 3 min, before returning to the cell incubator.

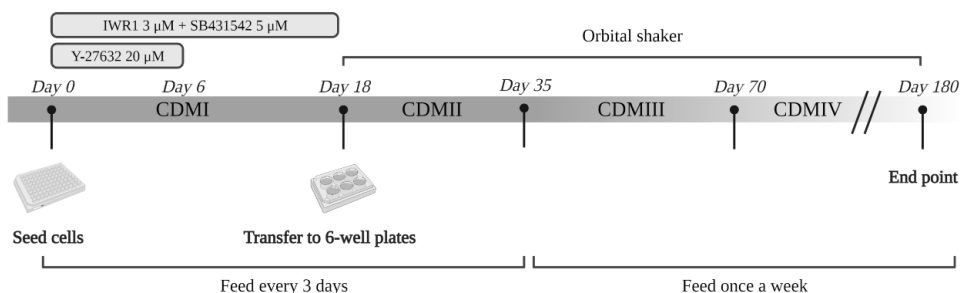


Figure 29. Schematic representation of the differentiation protocol used for the generation of human brain organoids, as described by Velasco et al²⁸⁶. CDM: cortical differentiation medium. Image created with Biorender.

Table 17. Recipe for preparation of cortical differentiation medium 1 (CDMI).

REAGENT	REFERENCE	STOCK CONCENTRATION	FINAL CONCENTRATION	VOLUME TO PREPARE 100ML
Glasgow-MEM	11710035 (Gibco)	100 %	77 %	77 mL
Knockout Serum Replacement	10828028 (Gibco)	100 %	20 %	20 mL
MEM Non-Essential Amino Acids	11140050 (Gibco)	100 X	1 X	1 mL
Penicillin/streptomycin	15140122 (Gibco)	10000 U/mL 10000 µg/mL	100 U/mL 100 µg/mL	1 mL
Sodium pyruvate	11360070 (Gibco)	100 mM	1 mM	1 mL
2-mercaptoethanol	21985023 (Gibco)	55 mM	0.1 mM	0.182 mL

On day 3, 100 µl of the same medium and supplements were added to each well. Every three days from day 6 to day 15, cells were fed by removing 80 µL of medium from each well and adding 100 µl of CDMI, only supplemented with IWR1 3 µM and SB431542 5 µM.

At day 18, aggregates were transferred to ultra-low attachment 6-well plates (Corning) and the medium was changed to CDMII (see Table 18). Between 5 and 6 aggregates were put on each well, and the plates were maintained from this day on in an orbital shaker (ThermoFisher) inside the cell incubator, with gentle agitation at 70 rpm. From day 18 to 33, medium was replaced every three days with fresh CDMII.

Table 18. Recipe for Preparation of cortical differentiation medium 2 (CDMII).

REAGENT	REFERENCE	STOCK CONCENTRATION	FINAL CONCENTRATION	VOLUME TO PREPARE 100ML
DMEM/F-12, HEPES	31330038 (Gibco)	100 %	97 %	97 mL
GlutaMAX	35050061 (Gibco)	100 X	1 X	1 mL
Chemically defined lipid concentrate	11905031 (Gibco)	100 X	1 X	1 mL
N-2 Supplement	17502048 (Gibco)	100 X	1 X	1 mL

On day 35, medium was replaced with CDMIII (Table 19) and cells were fed once a week until day 70. From day 70 on, the same feeding schedule was maintained, but replacing medium with CDMIV instead (Table 20).

Table 19. Preparation of cortical differentiation medium 3 (CDMIII).

REAGENT	REFERENCE	STOCK CONCENTRATION	FINAL CONCENTRATION	VOLUME TO PREPARE 100 ML
DMEM/F-12, HEPES	31330038 (Gibco)	100 %	86 %	86 mL
Hyclone Foetal Bovine Serum (FBS)	SH30070.03 (Cytiva)	100 %	10 %	10 mL
GlutaMAX	35050061 (Gibco)	100 X	1 X	1 mL
Matrigel Basement Membrane Matrix	356234 (Corning)	100 %	1 %	1 mL
Chemically defined lipid concentrate	11905031 (Gibco)	100 X	1 X	1 mL
N-2 Supplement	17502048 (Gibco)	100 X	1 X	1 mL
Heparin	H3149 (Sigma-Aldrich)	10 mg/mL	5 µg/mL	50 µL

Table 20. Preparation of cortical differentiation medium 4 (CDMIV).

REAGENT	REFERENCE	STOCK CONCENTRATION	FINAL CONCENTRATION	VOLUME TO PREPARE 100 ML
DMEM/F-12, HEPES	11330032 (Gibco)	100 %	82 %	82 mL
HyClone FBS	SH30070.03 (Cytiva)	100 %	10 %	10 mL
B-27 supplement, without serum	17504044 (Gibco)	50 X	1 X	2 mL
Matrigel Basement Membrane Matrix	356234 (Corning)	100 %	2 %	2 mL
GlutaMAX	35050061 (Gibco)	100 X	1 %	1 mL
N-2 Supplement	17502048 (Gibco)	100 X	1 %	1 mL
Chemically defined lipid concentrate	1190503 (Gibco)	100 X	1 X	1 mL
Penicillin/Streptomycin	15140122 (Gibco)	10000 U/mL 10000 µg/mL	100 U/mL 100 µg/mL	1 mL
Heparin	H3149 (Sigma-Aldrich)	10 mg/mL	5 µg/mL	50 µL

3.1.6. Morphometric analysis of neurons derived from hiPSCs.

The morphometric analysis of neurons derived from hiPSCs in monolayers was conducted as published by Mucci *et al.* with minor modifications¹⁷⁵. This protocol facilitates the measurement of morphometric parameters by transfecting neuron with a GFP-encoding vector. For this purpose, the desired efficiency of the transfection should be below 1 %, which allows the visualization of single neurons without overlapping signals from surrounding cells.

Briefly, on day 1, the neurons in monolayer of interest were fed with medium without antibiotics. The following day (day 2), transfection complexes were prepared by mixing in a 3:1 ratio the transfection reagent PEI with the GFP-plasmid of interest. The GFP-vector used in this protocol, pLenti CMV GFP Neo (657-2), was a gift from Eric Campeau & Paul Kaufman (Addgene plasmid #17447)²⁸⁷. The schematic representation of the plasmid can be found in Figure 30. Transfection complexes were prepared by mixing 8.34 µg PEI in 12.34 µL Opti-MEM (Gibco) with 2.78 µg of plasmid in 12.34 µL Opti-MEM (Gibco). After a 15-minute incubation at room temperature, the complexes were added dropwise to the cells and plates

coverslips were mounted with polyvinyl alcohol mounting medium with DABCO. Slides were stored overnight at room temperature, and then moved to 4 °C for long-term storage.

Table 21. Antibodies used for the characterization of differentiated cell types.

ANTIBODY NAME	TYPE	REFERENCE	DILUTION
Human/Mouse/Rat SOX2 Antibody	Primary	AF2018 (R&D Systems)	1:100
Purified anti-Pax-6 Antibody	Primary	901302 (BioLegend)	1:100
Anti-MAP2 antibody	Primary	ab5392 (abcam)	1:5000
TAOK1 antibody	Primary	26250-1-AP (Proteintech)	1:200
Anti-KIAA1107	Primary	HPA024216 (Atlas Antibodies)	1:1000
Anti-Ctip2 antibody	Primary	ab18465 (abcam)	1:300
Anti-SATB1 + SATB2 antibody	Primary	ab51502 (abcam)	1:100
Anti-GAD67 antibody	Primary	ab26116 (abcam)	1:100
Anti-Dlx2 antibody	Primary	ab272902 (abcam)	1:500
VGLUT1 Monoclonal Antibody (CL2754)	Primary	MA5-31373 (Invitrogen)	1:5000
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 568	Secondary	A10042 (Invitrogen)	1:500
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, FITC	Secondary	F2765 (Invitrogen)	1:500
Goat Anti-Rat IgG H&L (Alexa Fluor® 488)	Secondary	ab150157 (abcam)	1:500
Goat Anti-Mouse IgG H&L (Alexa Fluor® 488)	Secondary	ab150113 (abcam)	1:500
Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Secondary	A-11055 (Invitrogen)	1:500
Alexa Fluor 647-conjugated AffiniPure Donkey Anti-Chicken IgY H+L	Secondary	703-605-155 (Jackson ImmunoResearch)	1:500
Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 594	Secondary	A32742 (Invitrogen)	1:500

Images were taken with a Leica DM4 B Thunder Imager and the software Leica Application Suite v3.7.4.23463. All of them were processed and quantified using ImageJ.

3.1.7.2. Immunofluorescence of hiPSCs-derived brain organoids.

When the organoids reached the time points of interest, a few of them were fixed to perform immunostainings to detect the presence of specific-neuronal makers (previously indicated in Table 21).

3.1.7.2.1. Fixation of organoids.

To fix organoids, they were taken from the culture plate and placed using a wide-bore 1000 µl tip into a 15 mL tube. After removing the surrounding culture medium, they were washed once with cold PBS, and then submerged into 4 mL of cold fixing solution (4 % sucrose in 4 % PFA), in which they stayed overnight at 4 °C. The next morning, the fixing solution was removed, and the organoids were washed 3 times with cold PBS. After that, 4 mL of cold 30 % sucrose solution was added to the organoids, which were placed again at 4 °C. The next day, the organoids were ready for sectioning.

3.1.7.2.2. Organoid cryosectioning.

Organoids were sectioned with a Mycrom HM525 NX (Epredia). Samples were first taken from the 30 % sucrose solution and placed on a disposable base mold using a wide-bore 1000 µl tip. After removing the rests of preservation solution, OCT mounting media was added to the base mold, paying attention to cover the organoids without generating air bubbles. At this point, samples were frozen in the cryostat at – 56 °C and then quickly removed from the base mold and fixed on a cryostat cutting base.

At first, 40 µm slices were taken until the samples were visible. From that point forward, 18 µm slices were cut and fixed to a Superfrost plus adhesion microscope slide (Epredia). When all the samples were finally cut, the slides were placed at -80 °C for preservation.

3.1.7.2.3. Immunofluorescence of organoids.

First, slides were taken from -80 °C to the bench and were allowed to set at room temperature for a couple of min. Then, they were rinsed three times with PBS to eliminate the remaining OCT mounting media, after which they were incubated for 1 hour in blocking solution, prepared as stated previously on 3.1.7.1.2. Afterwards, slides were incubated with the primary antibodies diluted in blocking solution overnight at 4 °C.

The next day, slides were washed twice with 0.25 % TPBS and incubated for 10 min in blocking solution. Following this, slides were incubated at room temperature for 2 h with the secondary antibodies and 1 µg/mL Hoechst 33342 diluted in blocking solution. After that, slides were washed once with 0.25 % TPBS and twice with PBS. Finally, slides were submerged in milli-Q H₂O and coverslips mounted with polyvinyl alcohol mounting medium with DABCO DABCO. Slides were stored overnight at room temperature, and before moving them to 4 °C for long-term storage.

Some images were taken with an Olympus BX61 Fluorescence Microscope and the software cellSens Dimension (Olympus Corporation), while others were taken with a Leica DM4 B Thunder Imager and the software Leica Application Suite v3.7.4.23463. All of them were processed and quantified using ImageJ.

3.1.8. Proteomic analysis of hiPSCs-derived cortical neurons.

3.1.8.1. Protein extraction from hiPSCs-derived cortical neurons.

Cortical-like neuronal cultures differentiated from hiPSCs were processed to extract proteins using the RIPA Lysis Buffer System kit (Santa Cruz Biotechnology). Briefly, after washing the cells with PBS, up to 2×10^6 cells were resuspended in 0.2 mL of RIPA lysis buffer supplemented with 1:100 protease inhibitors cocktail, 2 mM PMSF and 1 mM sodium orthovanadate. Samples were incubated at 4 °C for an hour, and then centrifuged at the same temperature for 30 min at 13500 xg. Finally, the protein-containing supernatant was recovered and kept at – 80 °C until further processing.

3.1.8.2. Proteomic analysis.

Proteins obtained from neuronal-like cultures were identified through multidimensional liquid chromatography and mass spectrometry-based proteomic analysis. Sample processing was conducted as a collaboration with the Proteomics Platform (*Instituto de*

Investigación Sanitaria de Santiago-IDIS, located in the *Complejo Hospitalario Universitario de Santiago-CHUS*)) lead by Susana B. Bravo López, PhD. Briefly, protein samples were fragmented with trypsin and then processed with a TripleTOF 6600 quadrupole time-of-flight mass analyser instrument.

Preliminary data analysis was executed with the help of FunRich v3.1.3²⁸⁹ and Scaffold v5.3.1 (Proteome Software, Inc) software, along with STRING v12.0 webtool²⁹⁰ (String Consortium 2023).

3.1.9. Single cell RNA-seq of human brain organoids.

The scRNA-seq experiment was carried out following the steps on Figure 31. For this experiment, three batches of brain organoids were generated following the protocol described in 3.1.5.2 (biological replicates) for the two cell lines of interest: Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1*-i2 and its isogenic control Ctrl.BS-PBiPs37-Sv4F-1. At two different time points (day 48 and day 62), two organoids from each batch were selected for scRNA-seq (the experimental design is represented in Figure 32). In the following subheadings it is described how they were processed for that purpose.

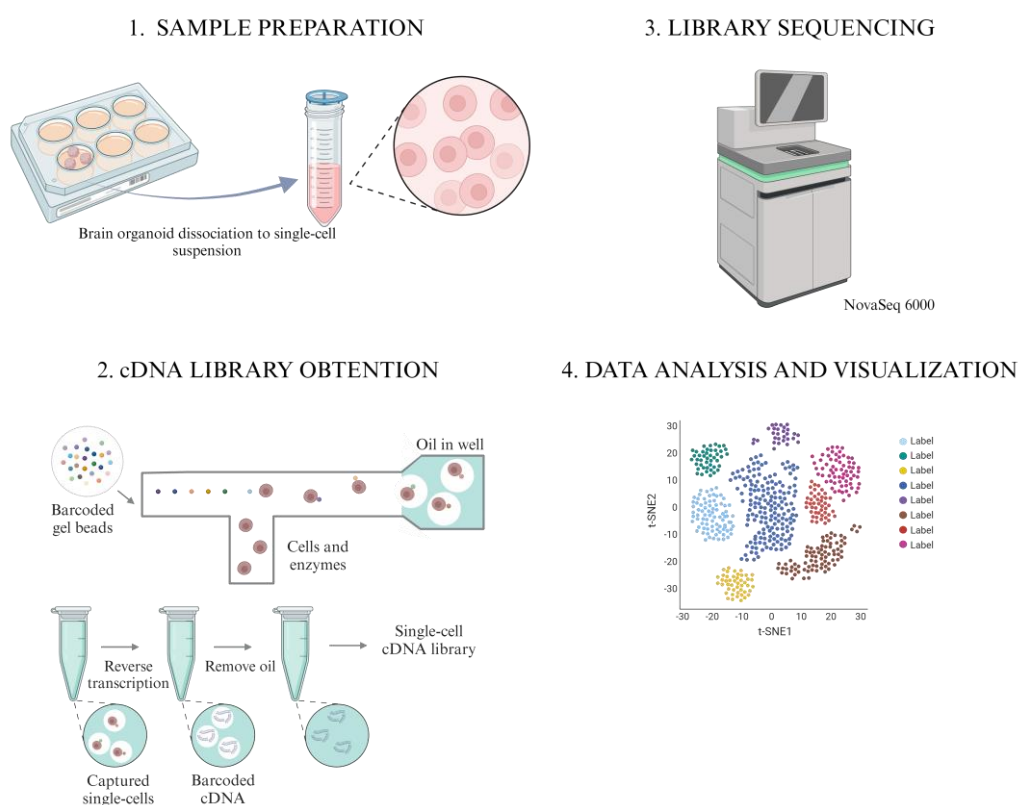


Figure 31. Workflow of a scRNA-seq experiment: from single cell obtention to data analysis. Image created with Biorender.

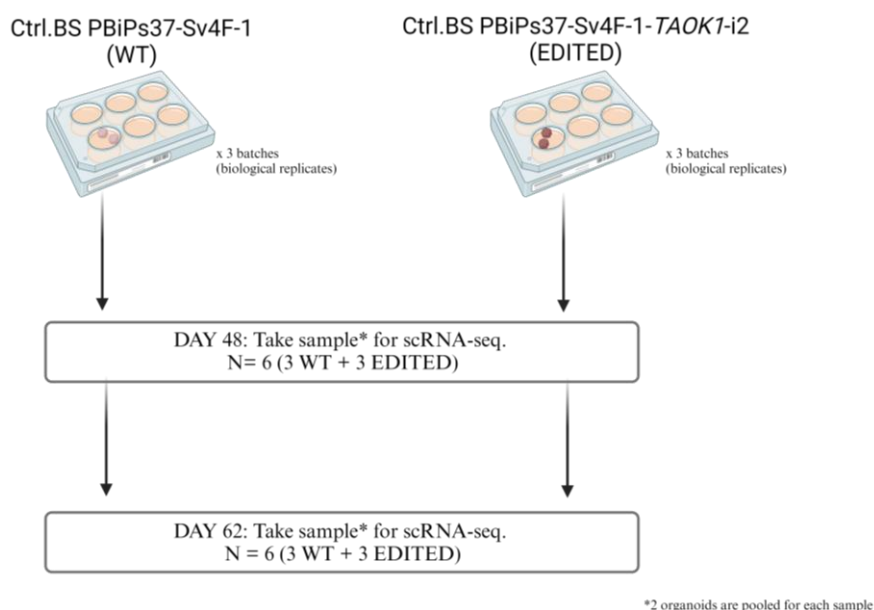


Figure 32. Experimental design of the scRNA-seq conducted on this thesis. Image created with Biorender.

3.1.9.1. Dissociation of organoids to single cells.

The first step to perform scRNA-seq consist in the dissociation of the organoids of interest to obtain single cells that can be analysed. To do so, we used a protocol based on the use of papain, as recommended by Velasco *et al.*²⁸⁶.

First, the components of the Papain dissociation system (Worthington Biochemical) were reconstituted following previously published indications²⁸⁶. Then, each organoid was transferred to a new 9.6 cm²-well filled with 2.5 mL of Papain and DNase I (10:1). There, each organoid was minced with a razor to until fragments smaller than 1 mm were obtained.

Table 22. Viability of the single-cell dissociated organoids used for scRNA-seq.

CELL LINE	TIME POINT	BATCH	VIABILITY AFTER SINGLE-CELL DISSOCIATION
Ctrl.BS-PBiPS37-Sv4F-1	Day 48	1	98.4 %
		2	97.2 %
		3	90.4 %
	Day 62	1	90.8 %
		2	93.8 %
		3	94.2 %
Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2	Day 48	1	98.7 %
		2	96.8 %
		3	99.7 %
	Day 62	1	94.4 %
		2	95.6 %
		3	94.4 %

The minced organoids were incubated at 37 °C for 30 min, after which they were resuspended a couple of times with a 1000 µl tip. Then, they were incubated 10 more min, after which the dissociated organoids were resuspended 10 times with a 10 mL serological pipette. The cell suspension was transferred to a 15 mL centrifuge tube in which it was allowed to settle for 3 min, to facilitate the precipitation of undissociated cell aggregates. The supernatant was mixed with 5 mL of Earle's Balanced Salt Solution (EBSS) and 3 mL of ovomucoid protease inhibitor solution in a new 15 mL centrifuge tube. After inverting the tube five times, it was

centrifuged at 300 xg for 7 min at room temperature. The supernatant was removed, and the cell pellet was resuspended in 250 μ L of 0.04 % BSA in PBS. At that point, the samples obtained from organoids of the same biological replicates were pooled together. Finally, the resuspended cells were filtered through a 37 μ m cell strainer to separate potential cell aggregates remaining in the solution. Cells were counted and their viability checked, since it is not recommended to perform scRNA-seq with a viability lower than 90 %. Viability counts for the samples used in the scRNA-seq experiment can be found in Table 22.

3.1.9.2. Library preparation and sequencing.

Library construction from single cells was conducted by a specialized service that is part of the GenVip research group (*Instituto de Investigación Sanitaria de Santiago-IDIS*, located in the *Complejo Hospitalario Universitario de Santiago-CHUS*). Briefly, 10000 single cells were separated into barcoded Gel Bead-In-EMulsions (GEMs) using a Chromium instrument and its compatible reagents (10x Genomics). In those GEMs, cDNA was generated by reverse transcription, adding a specific barcode for each individual cell. After a cleaning step, barcoded cDNAs were amplified by PCR and then used for library construction. Finally, library size was verified in a TapeStation High Sensitivity D1000 ScreenTape (Agilent) loading a dilution 1:10 of each sample (Figure 34).

The obtained Chromium Single Cell 3' Gene Expression (dual index) libraries (Figure 33) were sequenced in the *Fundación Pública Galega de Medicina Xenómica*. First, precise quantification was performed with a KAPA DNA Quantification kit (Roche). Afterwards, libraries were sequenced in a Novaseq 6000 instrument with a S2 flow cell (Illumina). The S2 flow cell is designed to pool up to 16 samples per run obtaining 20K reads per cell, but in this experiment the coverage is slightly increased since only 12 samples were pooled (all the samples on Figure 32 were pooled after the library generation for sequencing).

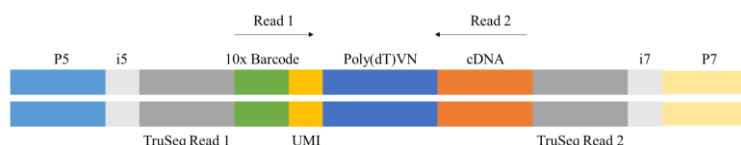


Figure 33. Structure of the Single Cell 3' Gene Expression libraries constructed for scRNAseq. The constructs start and end with Illumina-P5 and Illumina-P7 sequences, respectively. They also contain 10 bp sample index sequences (i5 and i7) and a poly(dT) site used to capture mRNAs in the reverse transcription step. TruSeq Read 1 and 2 are standard Illumina sequencing primer sites. TruSeq Read 1 is used to sequence the 10x Barcodes (16 bp) and unique molecule identifier (UMI, 12 bp), while TruSeq Read 2 is used to sequence the cDNA insert.

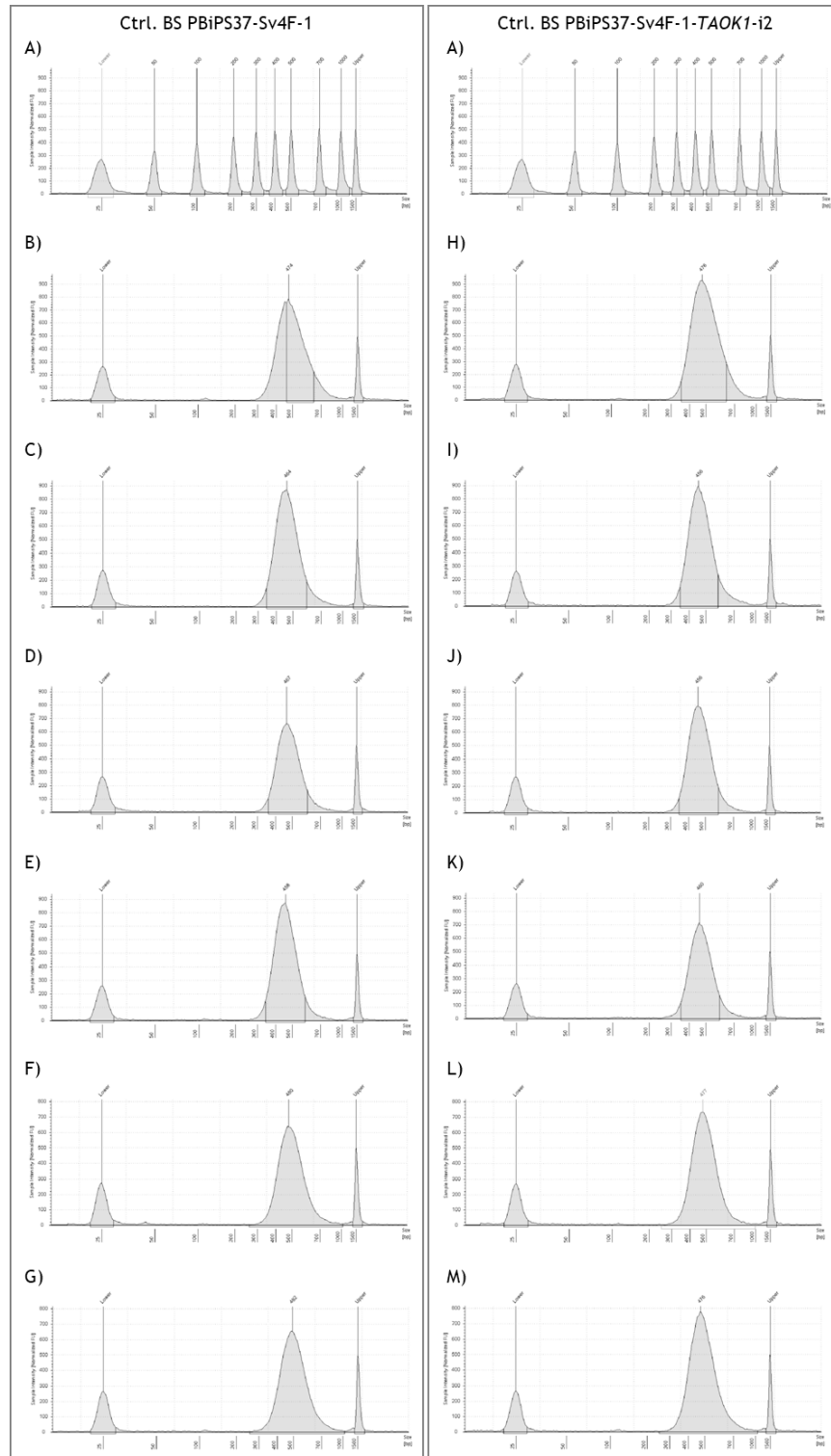


Figure 34. Library qualitative analysis prior to scRNA-seq (Agilent TapeStation High Sensitivity D1000 ScreenTape, TapeStation Analysis Software 4.1.1). Libraries were diluted 1:10 for this analysis. A) High Sensitivity D1000 Ladder. B) Ctrl.BS-PBiPS37-Sv4F-1, day 48 organoids, replicate 1. C) Ctrl.BS-PBiPS37-Sv4F-1, day 48 organoids, replicate 2. D) Ctrl. BS PBiPS37-Sv4F-1, day 48 organoids, replicate 3. E) Ctrl.BS-PBiPS37-Sv4F-1, day 62 organoids, replicate 1. F) Ctrl.BS-PBiPS37-Sv4F-1, day 62 organoids, replicate 2. G) Ctrl.BS-PBiPS37-Sv4F-1, day 62 organoids, replicate 3. H) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 48 organoids, replicate 1. I) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 48 organoids, replicate 2. J) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 48 organoids, replicate 3. K) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 62 organoids, replicate 1. L) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 62 organoids, replicate 2. M) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 62 organoids, replicate 3.

3.1.9.3. Single-cell RNA-seq data analysis.

Sequencing data was processed in the *Universidad de Málaga* by James Perkins, PhD. and Sergio Alías, PhD, who performed the quality control and cell-clustering using Seurat R package²⁹¹. Then, in our lab, cell types were assigned to each cluster based on their top 10 differentially expressed genes, which were compared to information available from brain organoids and human foetal brain accessed through UCSC Cell Browser^{292–296}. In addition, the expression of canonical markers of cell identity was investigated on each cluster^{119,297}.

Cluster composition was evaluated with a Chi-squared test ($\alpha = 0.005$), to identify deviations from the expected cell proportions.

3.2. PROTOCOLS USED FOR *IN VIVO* EXPERIMENTS IN ZEBRAFISH

The care, maintenance, use, and procedures described in this section were performed according to the current legislation (*Real Decreto 53/2013*) and with the approval of the Animal Care and Use Committee of the University of Santiago de Compostela following standard protocols (*Directiva 2010-63-UE*). Maintenance and procedures were performed in the animal facility EA-LU-003 (REGA authorization code ES270280346401), located in the Department of Zoology, Genetics and Anthropology (Veterinary Faculty, Lugo).

Procedures were performed by Sara Veiga Rúa, who has the capacitation for the functions A, B and C of animal experimentation in amphibians and fish.

3.2.1. Maintenance of zebrafish.

In the animal facility, zebrafish are kept in tanks with sweet water in a controlled and stable environment, with a temperature between 26-28 °C, water conductivity of 400-500 μ S and pH between 7-7.5. Tanks are connected by a close circuit with inversed osmosis, in which the water of the systems is constantly renewed. From the tanks, water flows to a canalization system composed of five different filters, before returning to the system. In order, these filters are wadding (mechanical filter), active carbon (chemical filter), mesh (mechanical filter), denitrifying bacteria (biological filter) and ultraviolet light.

Zebrafish are fed twice a day with dry food (GEMMA Micro) of a size according to the age of the individuals to avoid chocking.

3.2.1.1. Breeding.

Adult zebrafish individuals were stimulated to breed when fertilized eggs were needed. To do so, males and females were put together on special breeding tanks in the evening. These tanks had a rough bottom which mimics the natural environment that the animals use to lay eggs. In the morning, as the lights turned on in the animal facility, zebrafish mating began.

For some of the experiments described here, it was necessary to use eggs fertilized at the same time (synchronic eggs). For this, it was required to physically separate males and females in the breeding tanks in the evening. In the morning, the separator was removed as the lights turned on, which allowed individuals to mate. Around 10 min after the first eggs were laid, they were recovered and put on a Petri dish filled with inverse-osmosis water.

3.2.1.2. Anaesthesia and euthanasia.

Zebrafish were anesthetized with 0.002 % tricaine in inverse-osmosis water. Individual were maintained in the solution for a few min until the effects of anaesthesia were apparent, before performing procedures on the animals.

To euthanize embryos or adults, first the individuals were anesthetized, as indicated, after which an overdose of tricaine (0.2 % in inverse-osmosis water) was applied.

3.2.2. Generation of a morphant model of *fosl2* in zebrafish.

3.2.2.1. Microinjection of morpholinos.

Lyophilized morpholinos (MOs, GeneTools) were resuspended in sterile milli-Q H₂O to a final concentration of 1 mM. Before their use, morpholinos were incubated for 5 min at 60 °C to denaturalize secondary structures present in the molecules, and then they were placed on ice until their use.

Each zygote, at 0 hpf, was injected with 1-5 nL of a mixture of MOs and phenol red (0.2 %). The injection was conducted under a Leica binocular loupe with a microinjector (Narishige IM-31) connected to a nitrogen pump and suitable needles.

In this experiment, three conditions were included: non-injected control group (n = 112), group injected with MO-*fosl2* (n = 230), and group co-injected with MO-*fosl2* and MO-*p53* (n = 308). For the co-injected group, a mixture of both MOs was prepared in a 1:1 ratio. The sequences of the MOs that were used can be found in Table 23.

Table 23. Sequences of the morpholinos used in this thesis.

MORPHOLINO NAME	MORPHOLINO SEQUENCE (5'-3')
MO- <i>fosl2</i>	GCGCTGAGACACATCTGTGCATACC
MO- <i>p53</i>	AGAATTGATTTGCCGACCTCCTCT

After the injection, zygotes were incubated at 28.5 °C in inverse-osmosis water in Petri dishes.

3.2.2.2. Morphological analysis at 48 hpf.

The phenotype of the embryos was analysed 48 h after morpholine injection (48 hpf) to identify morphological alterations caused by the knockdown of the gene of interest.

After anaesthesia, photos of each embryo were taken using a AZ100 Multizoom microscope, associated to a DS-Ri1 camera (Nikon) with the help of the NIS-Elements software v4.13.04. Images were analysed using the software ImageJ to measure morphological parameters that are indicative of the development of zebrafish, which are: head area and diameter, eye area and diameter, as well as the full body length of the individuals.

Data was analysed in GraphPad Prism 8.0.2. First, the normality of the data was analysed applying the D'Agostino & Pearson test. Differences between normally distributed groups were assessed by the ANOVA test combined with Tukey's multiple comparison test. For non-normally distributed groups, the Kruskal-Wallis test was applied in combination with Dunn's multiple comparison test. For every statistical analysis, the considered p-value threshold was 0.05.

After taking the photos, embryos were euthanised. For each experimental condition, several pools of individuals were taken to perform RNA and DNA extractions. For RNA extraction, 3-

4 pools (15 embryos/pool) were taken, by placing the embryos in 30 μ L of RNAlater solution. For DNA extraction, 2-3 pools (7 embryos/pool) were prepared by including the embryos in 30 μ L of 100 % ethanol.



Figure 35. Morphometric analysis measurements conducted on zebrafish individuals: body length, cephalic size, and ocular size.

3.2.2.3. Behavioural analysis of the morphants using the light-dark test.

The behaviour of the larvae was analysed at 5 dpf, measuring the locomotion of individuals from the three experimental groups of interest (non-injected controls $n = 21$, MO-*fosl2* $n = 28$, MO-*fosl2* + MO-*p53* $n = 28$).

The locomotion of each individual was recorded for 60 min. During this time, larvae were exposed to three cycles of light and darkness, with a duration of 10 min each. Locomotion recordings were conducted in a Zebrafish observation chamber, coupled with the Zebrafish software (Viewpoint Life Sciences).

Data was exported and analysed in GraphPad Prism 8.0.2. First, the normality of the data was analysed applying the D'Agostino & Pearson test. Differences between normally distributed groups were assessed by the ANOVA test combined with Tukey's multiple comparison test. For non-normally distributed groups, the Kruskal-Wallis test was applied in combination with Dunn's multiple comparison test. For every statistical analysis, the considered p-value threshold was 0.05.

3.2.2.4. Alcian blue staining of cartilaginous tissue.

Alcian blue was used to stain cartilage in zebrafish larvae at 5 dpf (non-injected controls $n = 31$, MO-*fosl2* $n = 30$, MO-*fosl2* + MO-*p53* $n = 50$). First, larvae were euthanised, and then fixed overnight at room temperature in 4 % PFA in 24-well plates. The next day, the fixation solution was removed and three washes with PBS were performed. After that, larvae were incubated overnight with agitation in 0.1 mg/mL Alcian blue stain, prepared in ethanol:acetic acid (4:1). The following morning, the stain was removed, and larvae were treated with 0.6 % H_2O_2 in 1 % KOH for 2 h to depigment samples. After depigmentation, samples were washed once with PBS and then included in five solutions of 0.5 % KOH: 99 % glycerol, prepared with different ratios of each component (9:1, 7:3, 5:5, 3:7, 1:9, in order of use). Finally, samples were included in 99 % glycerol at room temperature before imaging.

Images were taken in a AZ100 Multizoom microscope, associated to a DS-Ri1 camera (Nikon) with the help of the NIS-Elements software v4.13.04.

The software ImageJ was used to analyse the photographs of the larvae, measuring the ceratohyal angle and palatoquadrate length.

Statistical analysis of the data was conducted using GraphPad Prism 8.0.2. First, the normality of the data was analysed applying the D'Agostino & Pearson test. Differences between normally distributed groups were assessed by the ANOVA test combined with Tukey's multiple comparison test. For non-normally distributed groups, the Kruskal-Wallis test was applied in combination with Dunn's multiple comparison test. For every statistical analysis, the considered p-value threshold was 0.05.

3.2.2.5. Knockdown verification by RT-PCR.

3.2.2.5.1. RNA extraction from zebrafish embryos.

RNA extraction from pools of zebrafish preserved in RNAlater was conducted using a protocol based on TRIzol/chloroform extraction (non-injected controls $n = 45$, *MO-fosl2* $n = 30$, *MO-fosl2 + MO-p53* $n = 60$). Briefly, embryos were washed once with cold PBS 1X before adding 250 μL of TRIzol to each pool, using a rod to help with the tissue disaggregation. After that, 100 μL of chloroform were added to each sample and everything was mixed using a vortex. Samples were then centrifuged at 13300 rpm for 20 min at 4 °C, after which the supernatant was recovered and mixed with isopropanol in a 1:1 ratio and 2 μL of glycogen 4 mg/mL. Following this step, samples were vortexed and incubated for 1 hour at – 80 °C. Once the incubation time was over, samples were centrifuged at 13300 rpm for 20 min at 4 °C. The pellet obtained on this step was washed with 500 μL of cold 70 % ethanol, after which the sample was centrifuged again for 5 min at 13300 rpm at 4 °C. Finally, the supernatant was discarded, and the pellet was resuspended in 15 μL of DEPC H₂O.

Then, to eliminate DNA from samples, each one was mixed with 5 μL of Buffer 10X and 1 μL of DNase I and incubated for 10 min at room temperature. Following the incubation, phenol-chloroform was added to the sample in a 1:1 ratio. Samples were centrifuged for 10 min at 13300 rpm at 4 °C, and the supernatant was recovered. This supernatant was mixed with cold 100 % ethanol in a 1:3 ratio and sodium acetate 3 M in a 10:1 ratio. Then, samples were placed in the – 80 °C to precipitate for 1 hour, after which they were centrifuged for 15 min at 13300 rpm at 4 °C. The pellet was washed with cold 70 % ethanol and centrifuged again for 10 min at 13300 rpm at 4 °C. Following this step, the supernatant was discarded, and the pellet resuspended in 20 μL of DEPC H₂O. Finally, RNA concentration was measured in a Nanodrop.

3.2.2.5.2. cDNA synthesis.

To reverse-transcribe RNA into cDNA, the AffinityScript Multiple Temperature cDNA Synthesis Kit (Agilent) was used.

Following the commercial instructions, 3 μL of random primers were added to each RNA sample (400 ng diluted in 12.7 μL H₂O). Then, reactions were incubated for 5 min at 65 °C, followed by a 10-minute incubation at 18 °C to allow the annealing of the primers. After that, each sample was mixed with the components indicated in Table 24 and reactions were incubated 10 min at 25 °C, followed by 1 hour at 50 °C and 15 min at 70 °C. Finally, samples were diluted to 10 ng/ μL and stored at -20 °C.

Table 24. cDNA synthesis reaction set-up.

REAGENT	VOLUME
Sample (RNA and random primers)	15.7 μ L
10 $\times$ AffinityScript RT Buffer	2 μ L
dNTPs (25 mM each)	0.8 μ L
RNase Block Ribonuclease Inhibitor (40 U/ μ L)	0.5 μ L
AffinityScript Multiple Temperature reverse-transcriptase	1 μ L

3.2.2.5.1. RT-PCR.

The design of primer pairs for RT-PCR was conducted with Primer Blast and Primer3. In Table 25 can be found the sequences and amplicon size of the designed primer pair for the verification of the knockdown of *fosl2* in zebrafish.

Table 25. Sequences and amplicon size of the RT-PCR primers.

GENE	FORWARD SEQUENCE (5->3)	REVERSE SEQUENCE (5->3)	AMPLICON SIZE
<i>fosl2</i>	TTACACCGGGACATACGACA	TACTGTGGGCTGAACCATCC	193 bp

PCR reactions of 15 μ L were prepared by mixing the following components: 8.6 μ L milli-Q H₂O, 1.5 μ L Buffer 10X, 1.05 μ L MgCl₂ 25mM, 0.75 μ L dNTPs 2mM, 1 μ L of each primer 10 μ M, 0.1 μ L Taq polymerase 5 U/ μ L and 1 μ L cDNA. Reactions were run in a thermocycler using the protocol indicated in Figure 36. Amplification was verified by gel electrophoresis in 1.5 % agarose gel. The gel was loaded with 4 μ L of each sample mixed with 1 μ L of bromophenol blue, including a well loaded with 100 bp DNA ladder. Electrophoresis was conducted for 35 min at 100 V, after which the gel was visualized in a transilluminator, and photographs were taken.

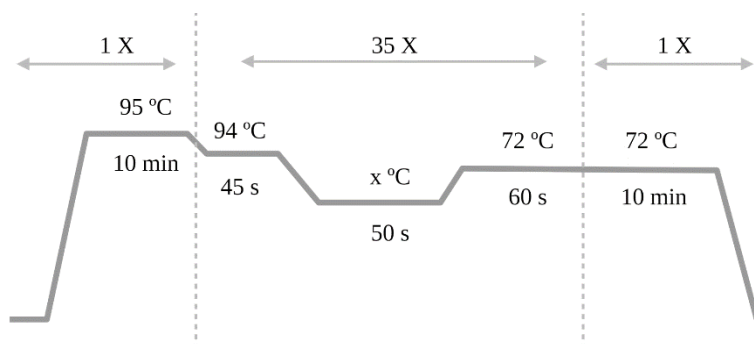


Figure 36. PCR conditions used in the amplification of CRISPR-Cas9 target sequences.

3.2.3. Generation of a mutant zebrafish line using CRISPR-Cas9

The first step to generate an animal model to study the function of a gene of interest is to verify that the chosen species has an orthologous gene that is highly conserved during evolution. Public databases (NCBI and ENSEMBL) were consulted to identify orthologous genes for the human genes of interest present in *Danio rerio*.

3.2.3.1. Guide design and synthesis.

As previously mentioned in the introduction, CRISPR-Cas9 guides have a region that will specifically bind the genome (crRNA) and a constant region that will bind the Cas9 protein (tracrRNA). Since the tracrRNA sequence does not change among guides, the only region that

needed to be designed for each experiment was the crRNA. Both sequences were ordered as DNA oligonucleotides, and then they were used as primers for a PCR reaction. The obtained PCR product was finally used as a template for the *in vitro* transcription of sgRNAs (Figure 37).

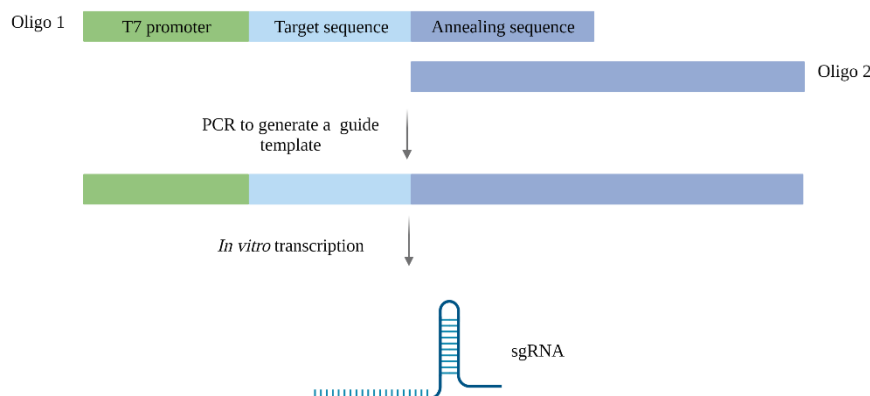


Figure 37. Schematic representation of the sgRNA synthesis for the knockout experiment in zebrafish. Image created with Biorender.com.

3.2.3.1.1. *In silico* knockout guide design.

The specific region of the guides was designed using the webtool CRISPRscan²⁹⁸ taking the genome assembly GRCz11 as a reference. Among the designed guides, only the ones that target all the gene's transcripts were considered. Guides with off-targets (up to 4 mismatches) were also discarded, and the rest of them were prioritized according to their expected efficiency. After applying these filters, two guides of 20 bp each were chosen to be used in the knockout experiment.

The designed guides were ordered as single-strand DNA oligonucleotides (primer gRNA). These sequences also included, in the 5' end, a T7 promoter and, in the 3' end, an annealing sequence that was complementary to the constant region of the guide (tracrRNA). Also, a single-strand DNA oligonucleotide for the tracrRNA was ordered (universal primer) (Table 26).

Table 26. Relevant sequences used for the knockout guide design in zebrafish. The sequences indicated in blue are complementary.

ELEMENT	SEQUENCE (5'-3')
T7 promoter	TAATACGACTCACTATA
Annealing sequence	GTTTTAGAGCTAGAA
DNA oligo corresponding to the tracrRNA fragment of the guide	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACG GACTAGCCTTATTTAACTTGCTATTTCTAGCTCTAAAC

3.2.3.1.2. Generation of a sgRNA template by PCR.

The sgRNA template was generated by running a PCR reaction prepared with 5 µL of High Fidelity Buffer 5X, 1.25 µL of dNTPs 2 mM, 1 µL of universal primer 10 µM, 1 µL of primer gRNA 10 µM, 0.25 µL of Phusion polymerase and 16.5 µL of H₂O, using the protocol indicated in the Figure 38.

The expected PCR product had a size of 117 bp, which was verified by gel electrophoresis in a 1 % agarose gel, prepared with 0.35 g agarose in 35 mL of 0.5X TBE buffer and 1.75 µL of SYBR Safe. The gel was loaded with 4 µL of each sample mixed with 1 µL of bromophenol blue, including a well loaded with 100 bp DNA ladder. Electrophoresis was conducted for 30

min at 100 V, after which the gel was visualized in a transilluminator, and photographs were taken.

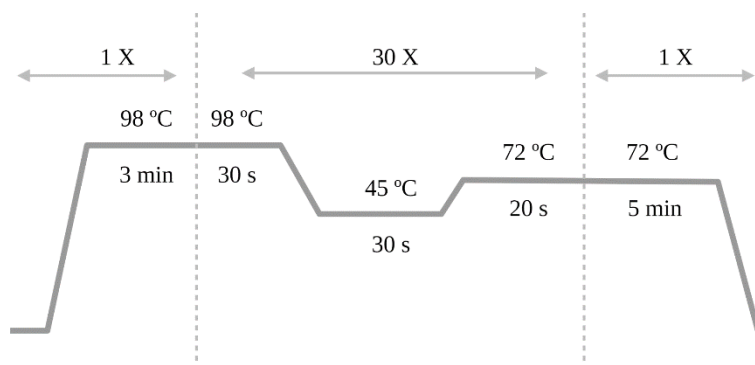


Figure 38. PCR program used to generate the sgRNA template for zebrafish knockout guides. Image created with Biorender.

3.2.3.1.3. Synthesis of sgRNA by *in vitro* transcription.

Once the sgRNA template was generated, it was transcribed *in vitro* with the MAXIscript T7 transcription kit (Invitrogen) to generate a sgRNA that could be used to inject zebrafish embryos. Briefly, the following reagents were added in order to a PCR tube: 3 μ L H₂O, 3 μ L PCR product (section 0), 1 μ L reaction buffer, 1 μ L enzyme mix, and 0.5 μ L of each rNTP 100 mM. After mixing everything, the reaction was incubated overnight at 37 °C.

The next day, 1.5 μ L of DNase turbo were added to the reaction and incubated for 30 min at 37 °C to digest the DNA sgRNA template. After that, 2 μ L of EDTA 0.5 M, 15 μ L of lithium chloride 5 M and 150 μ L of cold 100 % ethanol were added to the reaction, which was incubated at – 80 °C for 3 h and 45 min. Following the incubation, the reaction was centrifuged for 30 min at 13000 rpm at 4 °C. The obtained pellet was washed twice with 200 μ L of 70 % ethanol before resuspending it in 15 μ L of H₂O. Finally, RNA concentration was measured using a Nanodrop.

3.2.3.2. Genetic characterization of wild type zebrafish.

Before any CRISPR-Cas9 experiment is important to verify the sequence for the target region in the WT animals. This is relevant because different zebrafish strains might have polymorphisms that differ from the reference sequence available in public databases, which could impact the performance of the CRISPR-Cas9 guides.

For this reason, in addition to guides, primer pairs to sequence the target region of the guides were also designed with the help of the webtools Primer-BLAST (NCBI) and Primer3. These primer pairs were used to sequence several WT zebrafish to verify the sequence in the target region for CRISPR-Cas9.

3.2.3.2.1. DNA extraction from adult zebrafish.

DNA extraction from adult WT zebrafish was performed following a protocol based on the use of Chelex-100 and proteinase K. Each animal was anesthetized by keeping it in a 0.002 % tricaine in water solution. Once fish showed symptoms of unconsciousness, they were put on a Petri dish where a small fragment of caudal fin was cut with a scalpel. Immediately after, fish were put back in a tank with fresh water to recover from the anaesthesia.

The fragment of the caudal fin was put inside a 2 mL tube previously filled with 500 μ L of Chelex-100 (10 %) prewarmed at 60 $^{\circ}$ C. Then, samples were mixed and incubated in a water bath at 100 $^{\circ}$ C for 15 min. After the incubation, 15 μ L of proteinase K (20 mg/mL) were added per sample and, afterwards, incubated at 56 $^{\circ}$ C for 1 hour in an oven with constant rotation. Finally, samples were incubated for another 15 min in a water bath at 100 $^{\circ}$ C before storing them at -20 $^{\circ}$ C.

3.2.3.2.2. Amplification by PCR.

PCR reactions of 15 μ L were prepared by mixing the following components: 8.3 μ L milli-Q H₂O, 1.5 μ L Buffer 10X, 1.05 μ L MgCl₂ 25mM, 0.75 μ L dNTPs 2mM, 1 μ L of each primer 10 μ M, 0.1 μ L Taq polymerase 5 U/ μ L and 1.3 μ L DNA (extracted as described in 3.2.3.2.1). Reactions were run in a thermocycler using the protocol indicated in Figure 39. Amplification was verified by gel electrophoresis in 1 % agarose gels. The gel was loaded with 4 μ L of each sample mixed with 1 μ L of bromophenol blue, including a well loaded with 100 bp DNA ladder. Electrophoresis was conducted for 30 min at 100 V, after which the gel was visualized in a transilluminator, and photographs were taken.

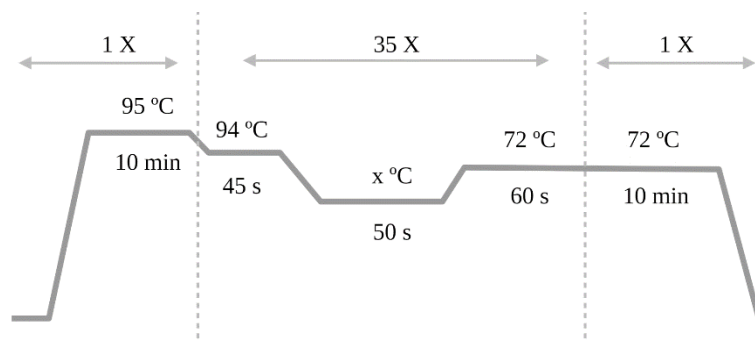


Figure 39. PCR conditions used in the amplification of CRISPR-Cas9 target sequences.

3.2.3.2.3. Sanger sequencing.

Before using the DNA amplified by PCR for sequencing, the samples were purified using the following protocol. First, 10 μ L of PCR product were mixed with 0.03 μ L of Exonuclease I 20 U/ μ L, 0.3 μ L of Thermosensitive Shrimp Alkaline Phosphatase (SAP) 1 U/ μ L and 1.67 μ L of milli-Q H₂O. After mixing all the reagents, samples were incubated for 1 hour at 60 $^{\circ}$ C followed by 15 min at 85 $^{\circ}$ C.

Next, samples were prepared to be sequenced by the Sanger method, using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and 96-well plates (4titude). For each reaction, the following components were mixed: 2 μ L buffer 5X, 0.5 μ L BigDye Terminator, 0.32 μ L of primer 10 μ M, 2 μ L of PCR purified product and 5.18 μ L of milli-Q H₂O. After mixing the reactions, they were incubated in a thermocycler programmed with the protocol indicated in the Figure 40.

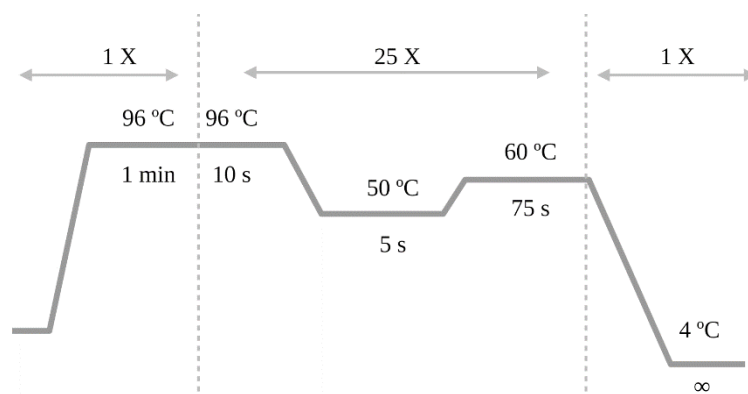


Figure 40. Thermocycler program used for Sanger sequencing with the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). Image created with Biorender.com.

Then, Sanger sequencing reactions were purified. To do so, samples were incubated at room temperature for 15 min in the dark with 10 μ L milli-Q H₂O, 5 μ L EDTA 125 mM and 60 μ L 100 % ethanol. Next, samples were centrifuged at 4000 rpm for 30 min at 20 °C and the supernatant was discarded. The plate containing the samples was centrifuged upside-down for 1 minute at 180 xg at 20 °C before adding 60 μ L of 70 % ethanol to each well. Then, the plate was centrifuged for the third time (15 min, 1650 xg, 4 °C) and the supernatant was discarded once more. The plate was centrifuged again upside-down (1 minute, 180 xg, 20 °C), after which each sample was resuspended in 10 μ L of formamide (FAM).

Finally, the sequencing plate was loaded into a 3730xl DNA Analyzer 96-capillary (Applied Biosystems). Sequencing results were analysed with the CodonCode Aligner software.

3.2.3.3. Microinjection of CRISPR-Cas9 components into zebrafish zygotes.

To generate a knockout mutant line for the gene of interest, it is necessary to deliver the components of the CRISPR-Cas9 system into synchronous embryos at 0 h post fecundation (hpf), which was achieved by microinjection. Each embryo was injected with 1-2 nL of a mixture containing 0.5 μ L of each sgRNA at 175 ng/ μ L, 0.5 μ L of True Cut Cas9 Protein v2 at 1 μ g/ μ L (Invitrogen), 1 μ L of KCl and 0.25 μ L of phenol red. The injection was done using special needles (1 mm external diameter, 0.58 mm internal microfilament, AM-Systems) and a microinjector (Narishige IM-31) connected to a nitrogen pump, with a balance pressure of 2,5-3 kPa and an injection pressure 30 kPa.

For this experiment, a total of 274 embryos were injected at 0 hpf. In the following subheadings, these individuals are going to be designated as F0 generation.

3.2.4. CRISPR-Cas9 editing efficiency analysis in individuals from the F0 generation.

Following the microinjection of the CRISPR-Cas9 components into zebrafish embryos, it's important to analyse its efficiency. Even though the editing system was injected in 0 hpf zygotes, that does not necessarily mean that the edition was conducted before the zygote underwent mitosis. Most likely, as a consequence of the edition, the individuals from the F0 generation have a mixture of both edited (in heterozygosis) and non-edited cells (wild type). An easy way to identify the presence of mutations in the individuals of the F0 generation is to perform an heteroduplex analysis in acrylamide gels. To do so, at 48 hpf were taken for analysis several pools of individuals: 4 F0 pools and 1 WT pool, each one composed of 5 embryos.

3.2.4.1. DNA extraction from zebrafish embryos.

DNA extraction from zebrafish embryos was performed following a protocol based on the use of Chelex-100 and proteinase K. Briefly, each individual was anesthetized by submersion in a solution of 0.002 % tricaine in water and, when they showed symptoms of unconsciousness, transferred to a 0.2 % tricaine solution to euthanize them. Afterwards, embryos were grouped into pools and transferred to a tube of 100 μ L of Chelex-100 (10 %) prewarmed at 60 °C. Then, samples were mixed and incubated in a water bath at 100 °C for 15 min. After the incubation, 3.1 μ L of proteinase K (20 mg/mL) was added per sample. Finally, samples were incubated at 56 °C for 1 hour followed by another 15 min at 100 °C, before storing them at -20 °C.

3.2.4.2. Amplification by PCR.

PCR reactions were performed as described previously in the subheading 3.2.3.2.2.

3.2.4.3. Heteroduplex analysis in acrylamide gels.

PCR products amplified from the pool of individuals of the F0 generation were analysed in acrylamide gels to detect the presence of heteroduplex, which would only be present if in the pool there were heterozygous individuals.

First, to prepare each acrylamide gel, the following reagents were mixed in order: 1.34 mL 29:1 acrylamide/bis-acrylamide (30 %), 1.6 mL TBE 5X, 5.03 mL distilled H₂O, 28 μ L APS 10 %, 6.67 μ L TEMED. Around 2 h later, after polymerization, 8 μ L of each sample were loaded into the acrylamide gels by mixing them with 1.6 μ L of bromophenol blue. Also, in the gels, a 100 bp DNA ladder was included.

After electrophoresis (40 min, 140 V, 0.02 A, 3 W), acrylamide gels were stained with a mixture of 2.5 μ L of SYBR Safe in 50 mL of distilled H₂O. Gels were incubated in this solution for 30 min in the dark with agitation. Then, gels were washed for 10 min in agitation in distilled H₂O, after which they were visualized in a transilluminator, and photographs were taken.

3.2.5. Genetic characterization of adult zebrafish from the F0 generation.

Once the zebrafish from the F0 generation reached adulthood, they were genetically characterized to try to identify which individuals had mutations in the *locus* of interest.

DNA extraction from adult zebrafish from the F0 generation was performed as described in 3.2.3.2.1. The only difference with regard of the steps previously indicated was that, in this case, each zebrafish was kept in a different compartment of a water tank to be able to identify to which one belongs each DNA sample. Those DNA samples were used as a template for PCR reactions (see 3.2.3.2.2) which were analysed in acrylamide gels to identify heteroduplex formation (see 3.2.4.3).

3.2.6. Generation of individuals from the F1 generation.

Once the zebrafish from the F0 generation with mutations in the *locus* of interest were identified (see 3.2.5), they were crossed with WT individuals (outcross). The objective of this step was to obtain a generation of zebrafish without mosaicism that had mutations in the *locus* of interest. The descendants from this cross are going to be referred from now on as F1 generation, which would be composed of a mixture of WT and heterozygous individuals.

3.2.6.1. Genetic characterization of adult zebrafish from the F1 generation.

Once the F1 generation reached adulthood, individuals were characterized as described previously for the animals of the F0 generation in the subheading 3.2.5 to differentiate the wild types from the heterozygous ones.

DNA samples from the heterozygous individuals were used to identify the exact mutations present in the locus of interest, which was achieved by cloning followed by sequencing, as well as fragment analysis.

3.2.6.2. Allele identification in the F1 population.

The alleles present in the F1 individuals were isolated using the cloning method previously described in the subheading 3.1.3.4. The obtained samples were sequenced following the protocol described in 3.2.3.2.3.

3.2.6.3. Fragment analysis.

Fragment analysis is a technique that determines the size of the DNA sequences present in a sample by comparing them with a size standard, which requires that both have a fluorescent mark to be detected. This technique was applied to identify the presence of insertions or deletions in the target region for the CRISPR-Cas9 system in zebrafish. By doing so, it is possible to genetically characterize mixed populations in a fast and easy way.

To do so, the region was amplified by PCR using a forward primer with a FAM marker attached to the 5' end (6-FAM, maximal signal emitted at 522 nm). The PCR reaction was performed using the reagents, concentrations and conditions described in 3.2.3.2.2.

In parallel, a mixture of 897 µl of FAM and 23 µl of the size standard GeneScan 500 LIZ dye was prepared. The size standard had 16 fragments (between 35-500 bp) marked with LIZ fluorophore (maximal signal emitted to 655 nm), which would be used as a reference to identify the size of the PCR fragment in the samples of interest. The FAM-marked PCR product was then prepared for fragment analysis in a 96-well plate, by mixing on each well 0.5 µl of PCR together with 9 µl of the previously prepared mixture. The plate was sealed and denaturalised for 3 min at 97 °C, after which it was loaded in a 3730xl DNA Analyzer 96-capillary (Applied Biosystems). Data was visualized using the program GeneMapper.

3.2.7. Generation of individuals from the F2 generation.

After the identification of the heterozygous mutants from F1 generation, crosses were established among them. By doing so, a generation zebrafish was obtained, which would be referred from now on as F2 generation. The F2 generation, as it was obtained by crossing heterozygous individual, in theory should be formed by 50 % of heterozygous mutants, 25 % of homozygous mutants and 25 % of WT zebrafish.

3.2.7.1. Characterization of individuals from the F2 generation.

The zebrafish individuals of the F2 generation were genetically characterized by fragment analysis as described in 3.2.6.3.

3.2.7.2. Morphological analysis.

The phenotype of the embryos was analysed 48 hpf (n = 37) and 96 hpf (n = 39) to identify morphological alterations caused by the knockout of the gene of interest. The analysis was carried out as previously described in the subheading 3.2.2.2.

After taking the photos, embryos were euthanised and cut in two sections using a scalpel. For each individual, the head section was preserved in 30 μ L of RNAlater solution for future RNA extraction, and the tail section was preserved in 30 μ L of 100 % ethanol for DNA extraction. Each tube was identified with the same code as the photos of the individual, to be able to trace the samples.

3.2.7.3. Behavioural analysis of the F2 generation using the light-dark test.

The behaviour of the larvae was analysed at 96 hpf, measuring the locomotion of individuals of the F2 generation (n = 39). The experiment was conducted as indicated in 3.2.2.3.

RESULTS

4. RESULTS

The results of this thesis will be presented in the same order of the proposed objectives. First, the generation and characterization of an *in vitro* model to study *BTBD8*'s function will be presented. Due to the lack of a preserved orthologous sequence in *Danio rerio*, an *in vivo* model to study *BTBD8*'s role was not created as a part of this thesis. Second, the generation and characterization of a knockdown zebrafish model to study the function of *fosl2* will be exposed. Third, the generation and characterization of *in vitro* and *in vivo* models to *TAOK1*'s function will be addressed.

4.1. GENERATION AND CHARACTERIZATION OF AN *IN VITRO* MODEL TO STUDY THE FUNCTION OF *BTBD8*.

As previously mentioned in the introduction section, a truncating mutation in *BTBD8* (NP_001363060.1:p.Ser937*) was found by Alonso-González and collaborators as potentially associated with ASD^{26,216}. Due to the lack of information available to support that relationship, one of the objectives of the present thesis was to develop an *in vitro* hiPSC-model to study the function of the gene and its role in neurodevelopment and its alterations.

To avoid potential misunderstandings in the following sections, it is relevant to clarify that the mutation linked to ASD was first identified in the GRCh37 human genome assembly. There, the mutation (C > G) was mapped to the gene *KIAA1107* (GRCh37, chr1:92632542-92650280) in the position chr1:92646050, leading to the generation of an early stop codon in the original 1354 aa protein (p.Ser499*).

That scenario changed slightly with the actualization of the human genome assembly GRCh38. In the GRCh38, the variant of interest (C>G) is in the position chr1: 92180490 (NP_001363060.1:p.Ser937*), which is included inside the gene *BTBD8*. *KIAA1107* (GRCg38, chr1: 92166985-92184723) is not considered an independent gene, but instead that genomic region forms part of the *BTBD8* (GRCh38, chr1: 92080305-92184725), previously located upstream *KIAA1107*.

4.1.1. Generation of a mutant hiPSC cell line for *BTBD8* using the CRISPR-Cas9 system.

The objective of this section is to describe the generation and characterization of a mutant hiPSCs model to study the function of *BTBD8* and its association with neurodevelopmental alterations.

4.1.1.1. CRISPR-Cas9 guide design and guide cloning into suitable vectors.

First, appropriate guides were designed to knockout the gene of interest were designed with CRISPR-ERA webtool²⁸⁰. Since this webtool works with the GRCh37 genome assembly, the selected target gene was *KIAA1107*, since that was the gene originally associated with ASD. As an output, the webtool generated a total of 280 potential guides to knockout the gene of

interest. Next, the most suitable guides to be expressed with the U6 promoter were selected, applying such filter in the same webtool.

Following that step, the resulting top 12 guides were analysed in the Synthego's Verify Guide Design²⁸² webtool, which uses the GRCh38, to verify the efficiency and suitability of the obtained guides (common exon targeting and minimal off-target effects). As a result, the best 3 guides according to the previously mentioned criterion were selected to test their editing efficiency *in vitro* (Figure 41).

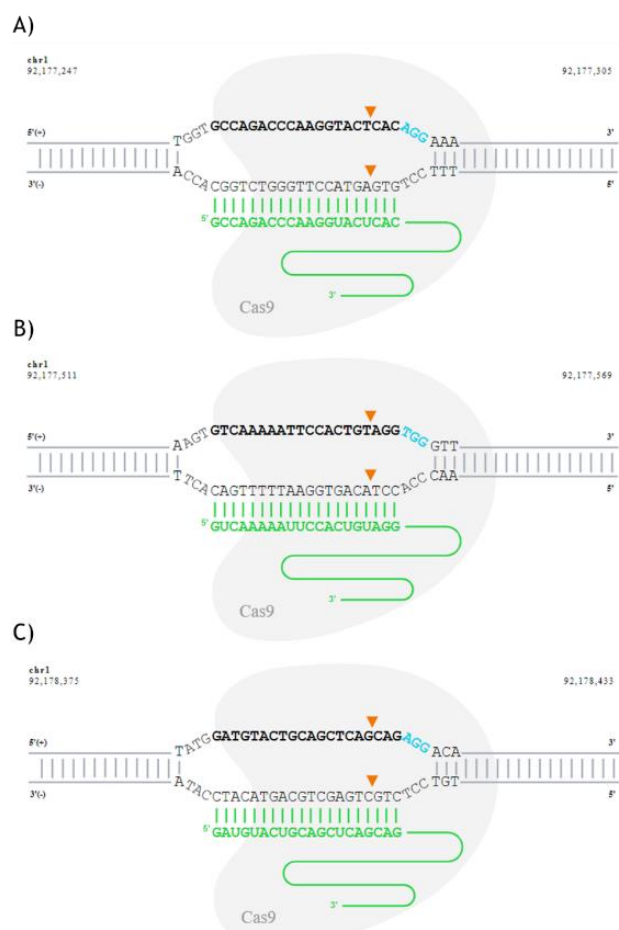
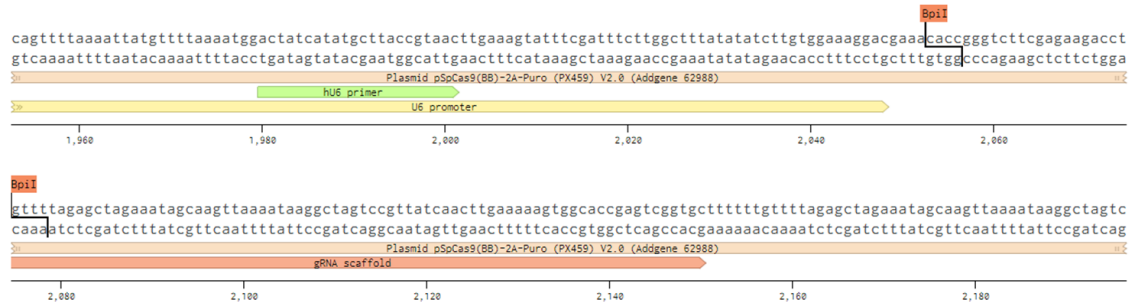


Figure 41. Schematic representation of the sgRNAs designed for the knockout of *BTBD8* on hiPSCs using the CRISPR-Cas9 system. A) *BTBD8*_sgRNA_1. B) *BTBD8*_sgRNA_2. C) *BTBD8*_sgRNA_3. Image generated by the author using Synthego's Verify Guide Design webtool v1.3²⁸².

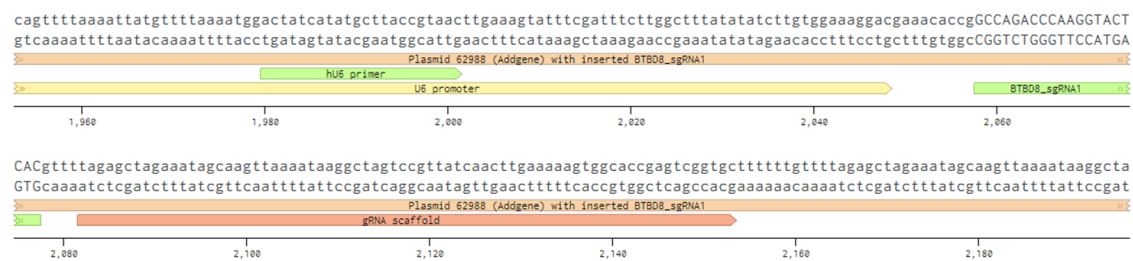
The next step was to insert the guides separately into the plasmid vector pSpCas9 (BB)-2A-Puro (Addgene #62988), which was conducted with the help of the restriction enzyme *Bpi*I. A prediction of the expected results was performed, as indicated in Figure 42. There, it can be observed the seamless insertion of each guide between the U6 promoter and the sgRNA scaffold (tracrRNA). The predicted sequences generated on that step were used as a reference to align the sequences of the cloned plasmids for each guide, to identify the ones correctly cloned. In Figure 43 can be observed that most of the analysed plasmids align perfectly with their predicted sequence, although some of them include small modifications that could disturb the functionality of the plasmid. In this case, the cloned plasmids selected for transfecting the CRISPR-Cas9 machinery into hiPSCs were *BTBD8*_sgRNA1_sample1,

*BTBD8*_sgRNA2_sample 4 and *BTBD8*_sgRNA3_sample 3, all of them represented in Figure 43.

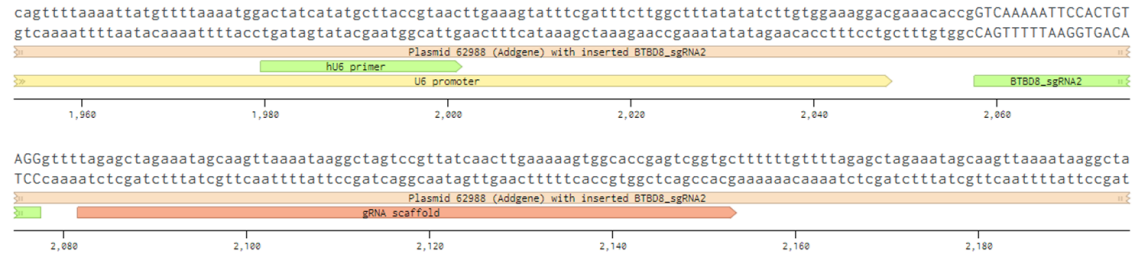
A)



B)



C)



D)

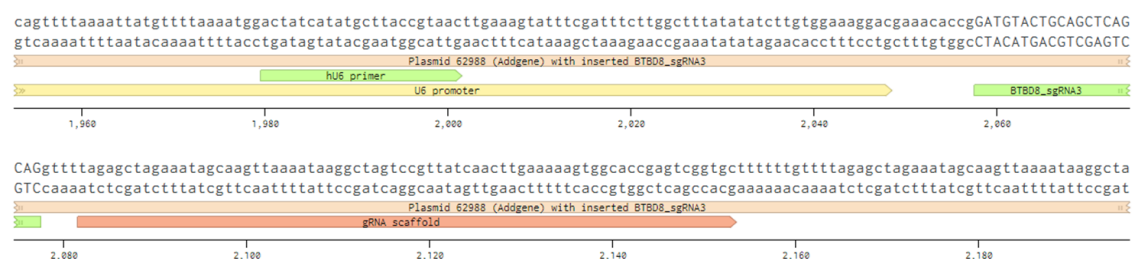


Figure 42. Representation of the plasmid pSpCas9 (BB)-2A-Puro (Addgene #62988) before and after the insertion of the guides to knockout *BTBD8*. A) Plasmid 62988 without modifications. B) Plasmid 62988 after the insertion of the *BTBD8*_sgRNA1. C) Plasmid 62988 after the insertion of the *BTBD8*_sgRNA2. D) Plasmid 62988 after the insertion of the *BTBD8*_sgRNA3.

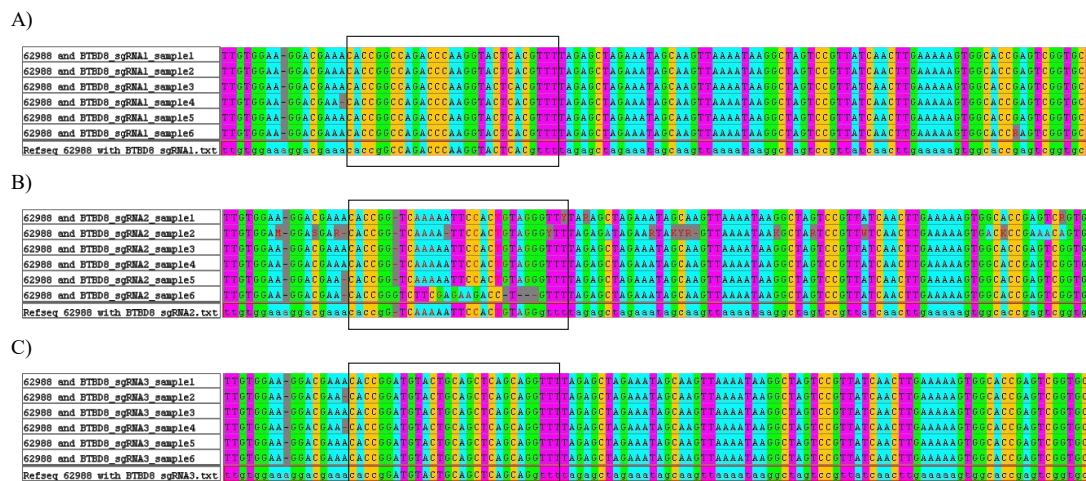


Figure 43. Verification of the correct insertion of the knockout guides for *BTBD8* in the plasmid pSpCas9 (BB)-2A-Puro (Addgene #62988) by Sanger sequencing. A) Plasmids cloned with *BTBD8_sgRNA1* (samples 1-6) aligned with the predicted sequence expected after the insertion (indicated as “Refseq” in the image). B) Plasmids cloned with *BTBD8_sgRNA1* (samples 1-6) aligned with the predicted sequence expected after the insertion (indicated as “Refseq” in the image). C) Plasmids cloned with *BTBD8_sgRNA1* (samples 1-6) aligned with the predicted sequence expected after the insertion (indicated as “Refseq” in the image). Black rectangles indicate the position of the inserted guide.

4.1.1.2. Puromycin kill curve for the cell line Ctrl.BS-PBiPS37-Sv4F-1.

Since the plasmid pSpCas9 (BB)-2A-Puro (Addgene #62988) contains a puromycin resistance gene, it was important to calculate the concentration needed to eliminate the non-resistant cells. To do so, a kill curve assay was conducted, which is represented in Figure 44.

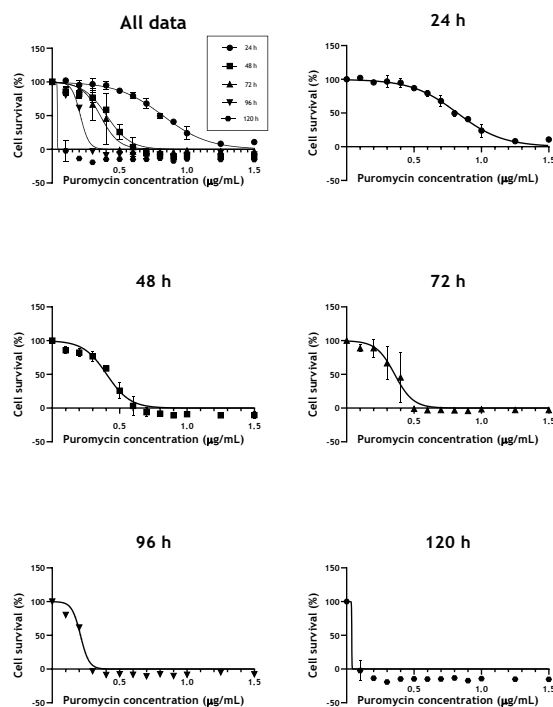


Figure 44. Puromycin kill curve for the cell line Ctrl.BS-PBiPS37-Sv4F-1. MTT viability assay performed after antibiotic treatment (24 h, 48 h, 72 h, 96 h, and 120 h). Data points represented as the mean $\pm$ SD.

As a result of this assay, it was determined that 1 $\mu\text{g/mL}$ of puromycin applied for 48 h was sufficient to eliminate non-resistant cells in the Ctrl.BS-PBiPS37-Sv4F-1 population.

4.1.1.3. Efficiency analysis of the knockout guides designed for *BTBD8*.

As stated earlier, three guides were designed to efficiently knockout *BTBD8*. Despite being theoretically optimal, a guide efficiency analysis was conducted to identify the one with the best performance. The method that was chosen for this purpose was the isolation of alleles in the pools of edited cells using bacterial cloning, followed by the sequencing of those fragments to identify the presence of mutations derived from the activity of the CRISPR-Cas9 machinery.

As can be observed in Figure 45 and Table 27, all the analysed alleles derived from the pool of edited cells with *BTBD8*_sgRNA1 showed mutations in the target region for the guide. However, only 33.33 % and a 16.67 % of the sequences were edited in the *BTBD8*_sgRNA2 (Figure 46, Table 28) and *BTBD8*_sgRNA3 (Figure 47, Table 29) pools of cells, respectively. These results support the selection of the pool of cells edited with *BTBD8*_sgRNA1 to derive a homogenous hiPSCs cell line knockout for *BTBD8*.



Figure 45. Alignment of the reference sequence of *BTBD8* (GRCh38), on top, with the alleles isolated by TOPO cloning from the bulk of edited cells. The target region of *BTBD8*_sgRNA1 is indicated in yellow.

Table 27. Sequence alterations identified by TOPO cloning of the cells treated with *BTBD8*_sgRNA1.

SEQUENCE BARCODE	TARGETED GENE	EXON	GUIDE	BACTERIAL CLONE	SEQUENCE ALTERATIONS
EF70647339	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA1	1	Deletion of 1 nt
EF70647340	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA1	2	Deletion of 11 nts
EF70647341	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA1	3	Deletion of 1 nt
EF70647342	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA1	4	Insertion of 1 nt
EF70647343	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA1	5	Deletion of 27 nts
EF70647344	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA1	6	Deletion of 1 nt



Figure 46. Alignment of the reference sequence of *BTBD8* (GRCh38), on top, with the alleles isolated by TOPO cloning from the bulk of edited cells. The target region of *BTBD8*_sgRNA2 is indicated in red.

Table 28. Sequence alterations identified by TOPO cloning of the cells treated with *BTBD8*_sgRNA2.

BARCODE	TARGETED GENE	EXON	GUIDE	BACTERIAL CLONE	SEQUENCE ALTERATIONS
EF70647345	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA2	1	None
EF70647346	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA2	2	None
EF70647347	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA2	3	None
EF70647348	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA2	4	None
EF70647349	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA2	5	Insertion of 1 nt
EF70647350	<i>BTBD8</i>	5	<i>BTBD8</i> _sgRNA2	6	Deletion of 6 nts and several base changes.



Figure 47. Alignment of the reference sequence of *BTBD8* (GRCh38), on top, with the alleles isolated by TOPO cloning from the bulk of edited cells. The target region of *BTBD8*_sgRNA3 is indicated in green.

Table 29. Sequence alterations identified by TOPO cloning of the cells treated with *BTBD8*_sgRNA3.

BARCODE	TARGETED GENE	EXON	GUIDE	BACTERIAL CLONE	SEQUENCE ALTERATIONS
EF70647333	<i>BTBD8</i>	7	<i>BTBD8</i> _sgRNA3	1	None
EF70647334	<i>BTBD8</i>	7	<i>BTBD8</i> _sgRNA3	2	None
EF70647335	<i>BTBD8</i>	7	<i>BTBD8</i> _sgRNA3	3	Deletion of 11 nts
EF70647336	<i>BTBD8</i>	7	<i>BTBD8</i> _sgRNA3	4	None
EF70647337	<i>BTBD8</i>	7	<i>BTBD8</i> _sgRNA3	5	None
EF70647338	<i>BTBD8</i>	7	<i>BTBD8</i> _sgRNA3	6	None

4.1.2. Characterization of the generated clones.

The pool of cells edited with *BTBD8*_sgRNA1 were seeded at low density, with the objective of growing single-cell derived colonies that could be isolated. In the following subheadings it is presented the process characterization of the clones obtained as a result.

4.1.2.1. Allele identification.

To begin with the characterization process, the presence of mutations in *BTBD8* was addressed. Here, 44 clones were isolated from the pool of edited cells with *BTBD8*_sgRNA1, which were sequenced to find signs of genomic editing events in the target region of the guide. One of the 44 clones (clone 5, Figure 48) showed a clean wild type sequence, with no signs of genetic modification in the region of interest, while the remaining 39 had alignment problems, thus suggesting potential indels affecting the area. In addition, four clones had a clear Sanger sequence with mutations apparently in homozygosity (Figure 49), but this observation could also

have been made if the detected mutation affected one allele, while the other had a large deletion, leading to increased levels of genomic instability. To avoid this possibility, those clones were not further used on this thesis.

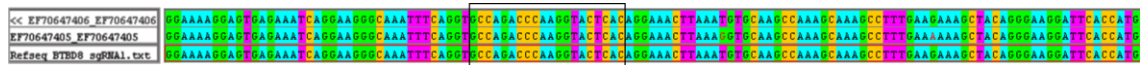


Figure 48. Clone 5 isolated from the bulk of cells edited with *BTBD8*_sgRNA1. Alignment of the reference sequence of *BTBD8* (GRCh38) with the Sanger sequences obtained from the clone of interest.

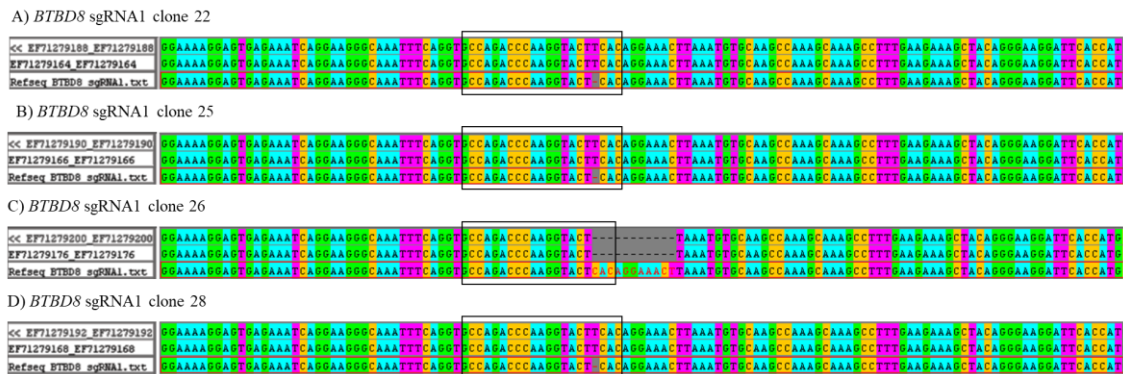


Figure 49. Alignment of the reference sequence of *BTBD8* (GRCh38) with the Sanger sequences obtained from several clones of interest isolated from the bulk of cells edited with *BTBD8*_sgRNA1. A) Clone 22 with an apparent insertion of 1 nt in homozygosis. B) Clone 25 with an apparent insertion of 1 nt in homozygosis. C) Clone 26 with an apparent deletion of 11 nts in homozygosis. D) Clone 28 with an apparent insertion of 1 nt in homozygosis.

Next, bacterial cloning was applied to 10 potentially edited clones to precisely identify the allele composition of the cell population. In clones 1, 2, 3, 4, 6, 7, 8, and 38, more than two alleles were identified for each clone, which suggests that they were mixed populations derived from different edited cells, and thus they were discarded from further analysis.

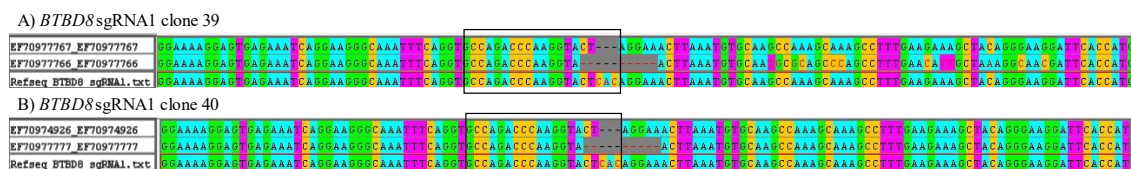


Figure 50. Alignment with the reference sequence of *BTBD8* (GRCh38) with the Sanger sequences obtained from several clones of interest after TOPO cloning. A) Alleles present in the clone 39 isolated from the bulk of cells edited with *BTBD8*_sgRNA1. B) Alleles present in the clone 40 isolated from the bulk of cells edited with *BTBD8*_sgRNA1.

However, clones 39 and 40 presented a clear heterozygous mutation in *BTBD8*, which was identical in both (Figure 50). One allele had a stop-gain deletion of 10 nts (NC_000001.11:g.92177279_92177288delCTCACAGGAA), while the other had a deletion of 3 nts (NC_000001.11:g.92177281_92177283delCAC). The 10 nts frameshift mutation is expected to generate a 696 aa protein (NP_001363060.1:p.Leu696Thrfs*1), while the 3 nt mutation is predicted to encode a 1791 aa protein (NP_001363060.1:p.Thr697del) with only one amino acid missing without frameshift alterations associated. As can be seen in Figure 51, the 10 nt deletion generates a shorter *BTBD8* protein, with abundant disordered regions, which presumably alter the function of the protein. The effects of the 3 nt deletion are less clear. One possibility is that the deletion does not have a major impact on the function of the protein, considering that the alteration affects a disordered region of the peptide. Nevertheless,

comparing the predicted 3D structures of the wild type BTBD8 protein (NP_001363060.1) and the mutated NP_001363060.1:p.Thr697del (Figure 11B and Figure 51A), differences are apparent, which could be associated to an altered function of the protein.

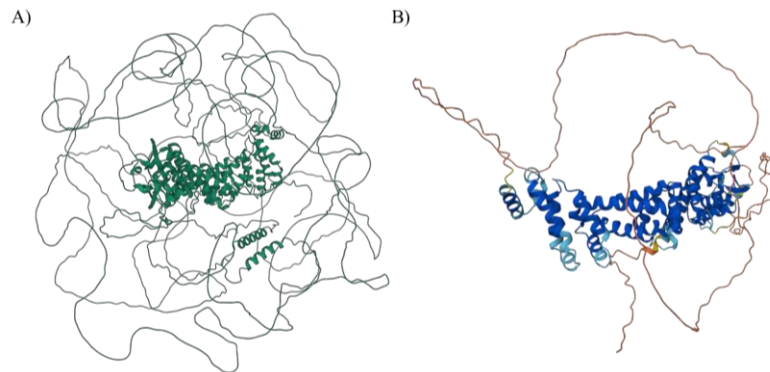


Figure 51. Predicted structures of the proteins generated by the mutant *BTBD8* alleles. A) Protein NP_001363060.1:p.Thr697del, encoded by the mutant *BTBD8* allele lacking 3 nucleotides (NC_000001.11:g.92177281_92177283delCAC). B) Protein NP_001363060.1:p.Leu696Thrfs*1, encoded by the mutant *BTBD8* allele lacking 10 nucleotides (NC_000001.11:g.92177279_92177288delCTCACAGGAA). Image generated using AlphaFold Monomer v2.0 pipeline^{214,215}.

After evaluating these results, *BTBD8*_sgRNA1_clone 39 was selected to be further characterized, since it is expected to show haploinsufficiency for the gene of interest, which was also detected in ASD patients, as mentioned in the introduction. In addition, the hiPSCs clone was re-named as Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2, following the guidelines of the *Banco Nacional de Líneas Celulares* and the *Instituto de Salud Carlos III*²⁸⁵, which will be used from now on to refer to that cell line.

4.1.2.2. Karyotype analysis of the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2.

The karyotype of the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2 was analysed with the objective of detecting potential chromosomal abnormalities. A total of 21 metaphases were scrutinized, with none of them displaying numerical or structural abnormalities, including gaps and breaks, under the resolution of the karyotype analysis. One of the analysed karyotypes can be found in Figure 52.

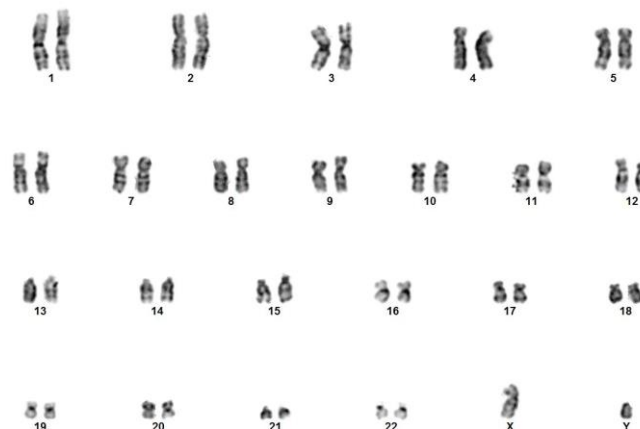


Figure 52. Karyotype of the cell line Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2.

4.1.2.3. Pluripotency marker expression in the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2.

To further continue the characterization of the edited hiPSC line Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2, the expression of pluripotency markers *NANOG*, *POU5F1* and *SOX2* was interrogated at the mRNA level.

As a first step, primer pairs were designed to analyse the expression of the previously mentioned markers, along with the housekeeping gene *GAPDH*, using RT-qPCR with SYBR Green chemistry. Since they were manually designed, it was also necessary to test their efficiency to ensure the reliability of the obtained results, which was analysed by doing a standard curve with 1/10 cDNA dilutions (Figure 53). As can be seen in Figure 53, the efficiency of all the designed primer pairs falls into the 90 – 110 %, which means that they have an optimal efficiency for their use in RT-qPCR assays.

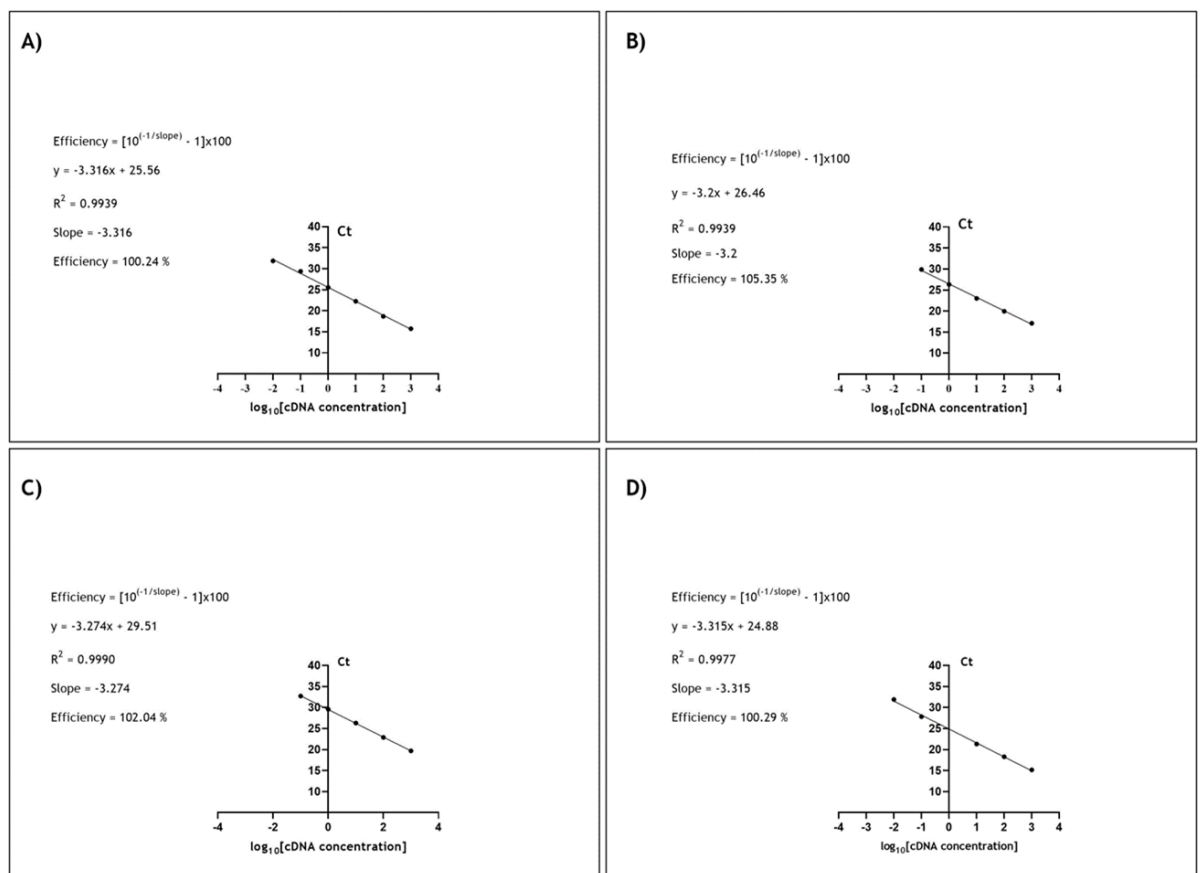


Figure 53. Standard curves for the efficiency analysis of RT-qPCR primer pairs. A) *GAPDH*. B) *SOX2*. C) *NANOG*. D) *POU5F1*.

After performing this verification step, those primer pairs were used to interrogate the expression of the pluripotency markers *NANOG*, *POU5F1*, and *SOX2* in the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2, in comparison with the isogenic control Ctrl.BS-PBiPs37-Sv4F-1. As can be observed in Figure 54, no significant differences were found for none of the pluripotent cell markers analysed.

In addition, the expression of pluripotency markers at the protein level was also studied. Using immunofluorescence, it was possible to detect the expression of the stem cell markers

NANOG, SOX2, and TRA-1-81 (Figure 55, Figure 56, Figure 57) in the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1*-i2, further confirming their pluripotency.

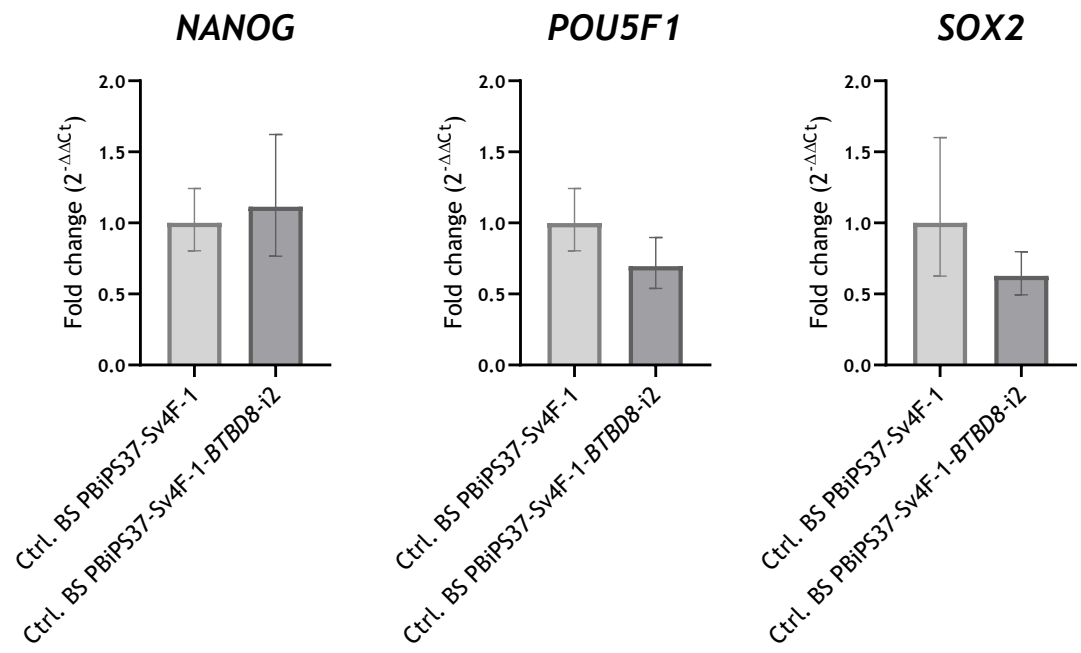


Figure 54. RT-qPCR for pluripotency markers in the cell line Ctrl.BS-PBiPs37_Sv4F-1_ *BTBD8*-i2 compared to the isogenic wild type cell line Ctrl.BS-PBiPs37_Sv4F-1. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

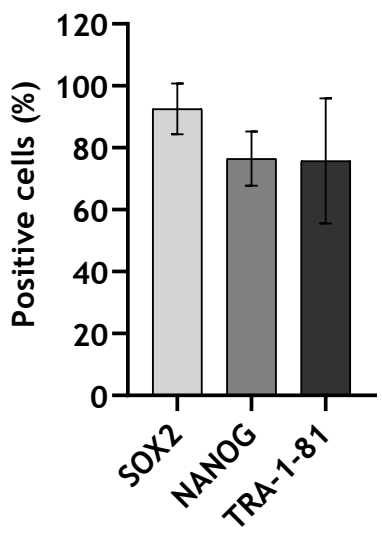


Figure 55. Quantification of immunofluorescence images for the pluripotency markers SOX2, NANOG, and TRA-1-81. Data represented as mean ± SD.

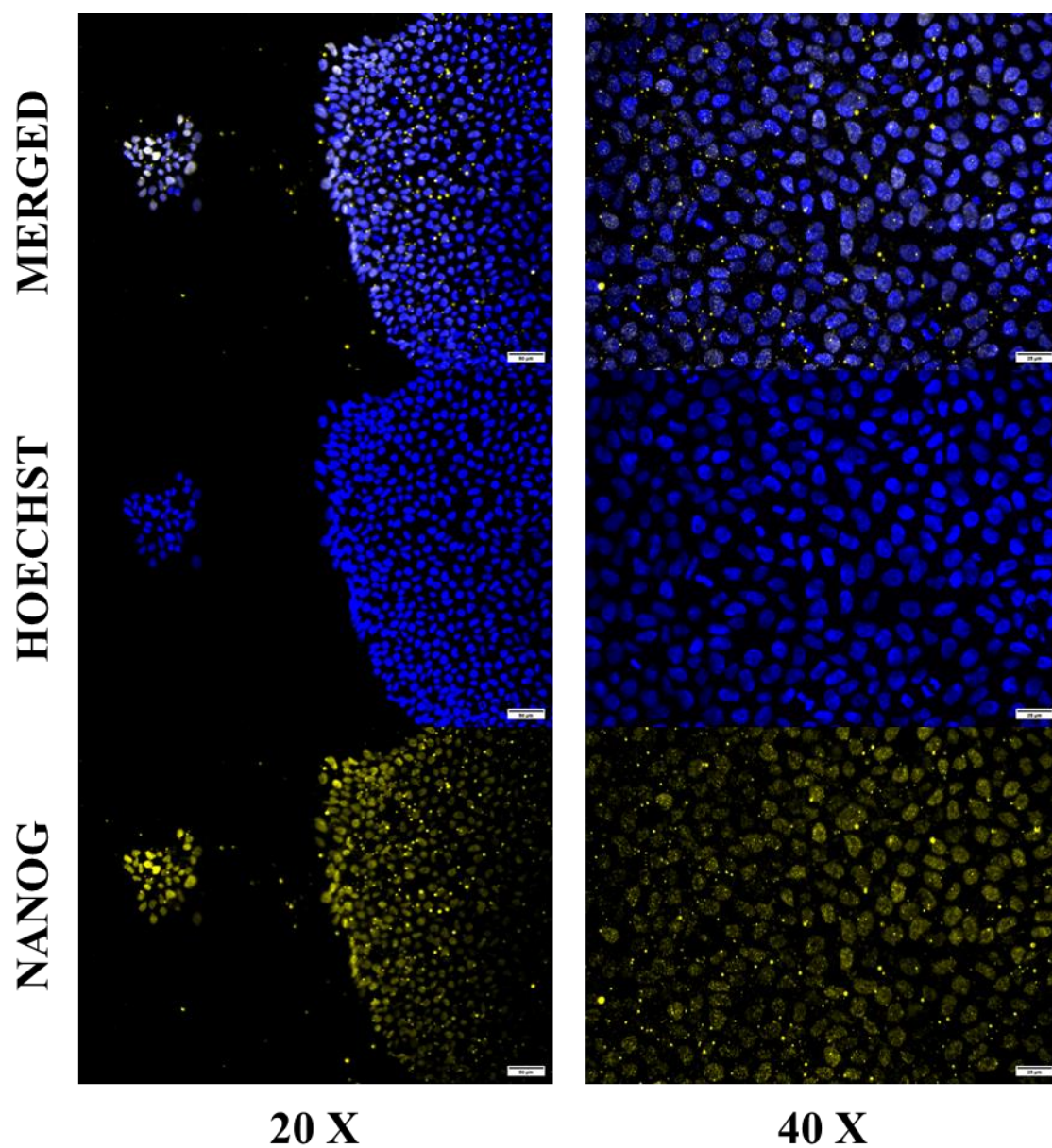


Figure 56. Immunofluorescence of the pluripotency marker NANOG in the cell line Ctrl.BS-PBiPs37-Sv4F-1-BTBD8-i2. Nuclear staining was performed with Hoechst.

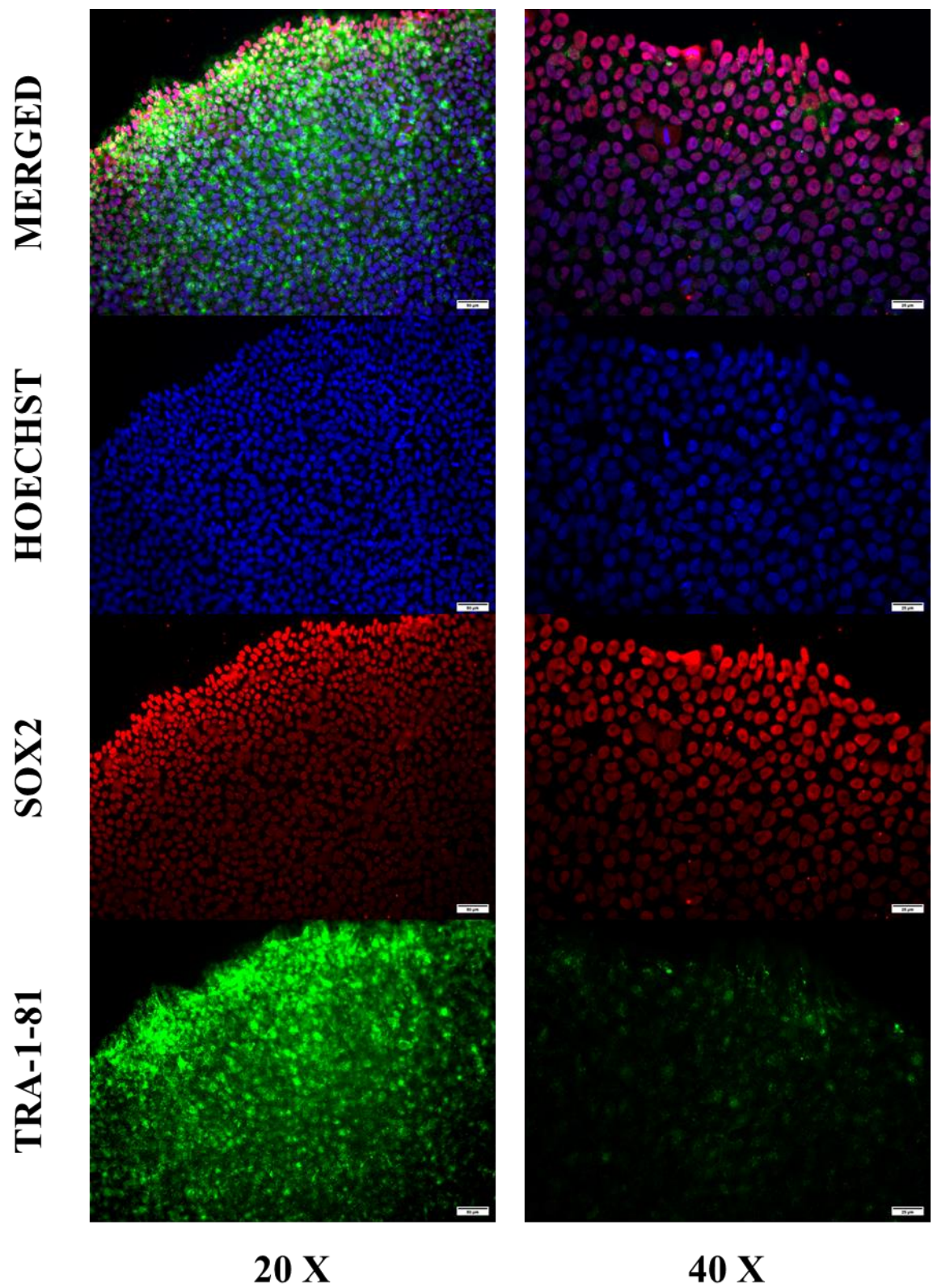


Figure 57. Immunofluorescence of the pluripotency markers SOX2 and TRA-1-81 in the cell line Ctrl.BS-PBiPs37-Sv4F-1-BTBD8-i2. Nuclear staining was performed with Hoechst.

4.1.2.4. Trilineage differentiation of Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2 edited cells.

The characterization of the edited hiPSC line Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2 continued with the evaluation of its pluripotency by conducting a trilineage differentiation assay. Pluripotency was assessed by inducing differentiation of hiPSCs into the three germ layers, including ectoderm, mesoderm, and endoderm.

Brightfield images taken at the start and end of the differentiation protocol for each germ layer are displayed in Figure 58. Visually, there are no striking differences between the *BTBD8*-mutant cells and their isogenic WT control for neither germ-layer.

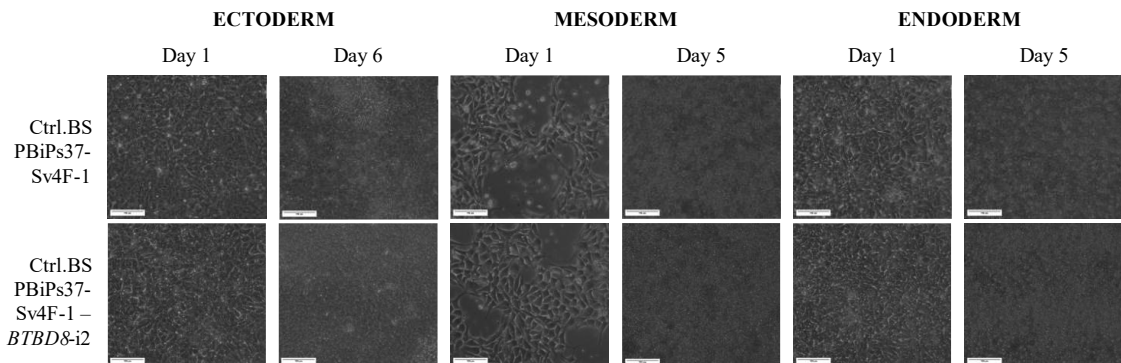


Figure 58. Brightfield images of Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2 edited cells and their isogenic control (Ctrl.BS-PBiPs37-Sv4F-1) taken during the three germ-layer differentiation protocol.

In addition to visually inspecting the differentiated cultures, RT-qPCR was used to determine if they expressed specific markers for each germ-layer. As a first step, primer pairs were designed to analyse the expression of *NES* (ectoderm), *NCAM1* (mesoderm) and *AFP* (endoderm) and their efficiency was tested with a standard curve. As can be seen in Figure 59, the efficiency of all the designed primer pairs falls into the 90 – 110 % range, thus having an optimal efficiency for their use in RT-qPCR assays. The standard curve for the primers used to amplify the housekeeping gene *GAPDH* has been previously presented in Figure 53.

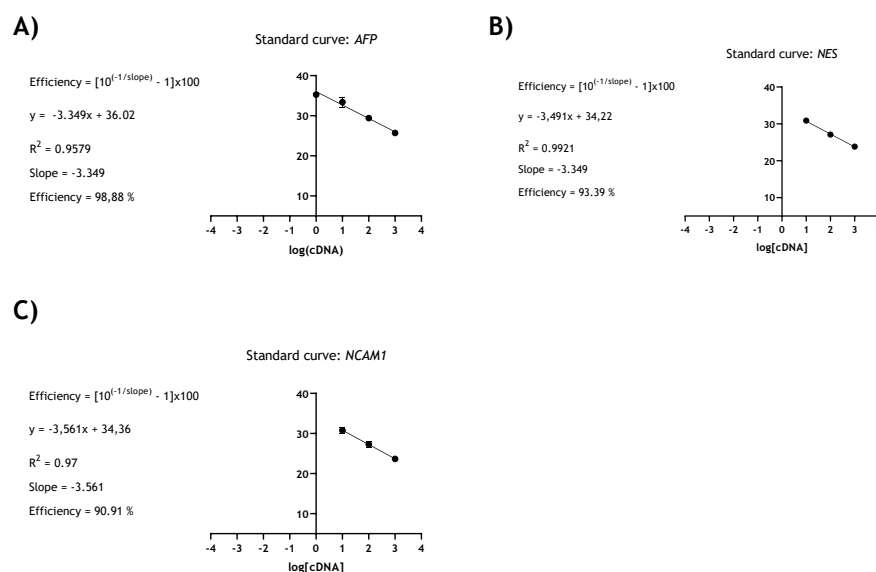


Figure 59. Standard curves for the efficiency analysis of RT-qPCR primer pairs. A) *AFP*. B) *NES*. C) *NCAM1*.

Next, the selected primers were used to interrogate the expression of each germ-layer specific marker in differentiated cultures of Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2 and its isogenic control. As can be observed in Figure 59, differentiated cells from both lines expressed the corresponding germ-layer markers: *NCAM1* expression was tested in mesoderm differentiated cells, *NES* expression was tested in ectoderm differentiated cells, and *AFP* expression was tested in endoderm differentiated cells.

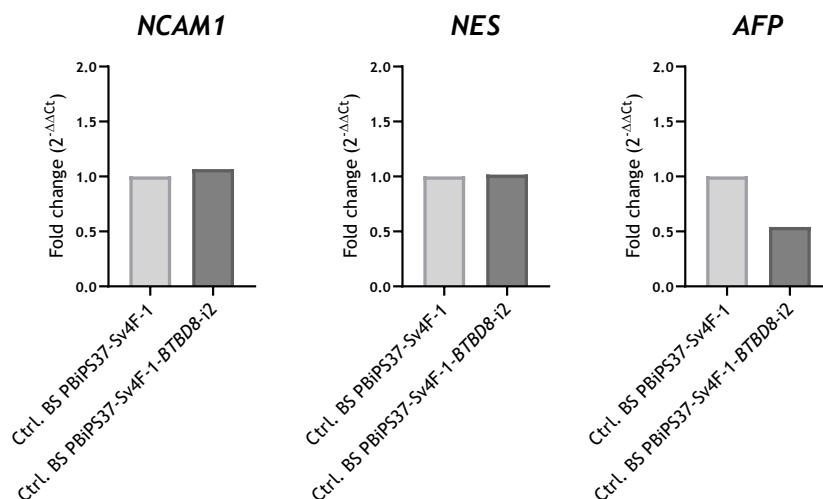


Figure 60. RT-qPCR for germ-layer markers in the cell line Ctrl.BS-PBiPs37_Sv4F-1-*BTBD8*-i2 comparison to the isogenic wild type cell line Ctrl.BS-PBiPs37_Sv4F-1. *NCAM1* expression was tested in mesoderm differentiated cells. *NES* expression was tested in ectoderm differentiated cells. *AFP* expression was tested in endoderm differentiated cells.

4.1.2.5. Off-target assessment of Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2 edited cells.

Table 30. Predicted off-target regions for *BTBD8*_sgRNA1 identified using the Cas-OFFinder⁶⁸ and Synthego's Verify Guide Desing²⁸² webtools. Mismatches between the off-target region and the sgRNA sequence are indicated in lowercase.

CODE	MISMATCH	OFF-TARGET SEQUENCE	CUT SITE	GENE
OFF_B_1	3	GgCAGACCCAAtGTAA TCACAGG	chr20:53,350,085	<i>TSHZ2</i>
OFF_B_2	3	GgCAGGCCCAAGGTAA TCACAGG	chr20:56,588,830	<i>LINC01716</i>
OFF_B_3	3	GCCAGACTCagGGTACcCACTGG	chr2:82,031,248	-
OFF_B_4	3	GCCAGACTCagGGTACcCACTGG	chr2:82,023,212	-
OFF_B_5	3	GCcGACaCAAGcTACTCACGGG	chr2:126,231,390	<i>ENSG00000290113</i>
OFF_B_6	3	GCCAGGcTCAAGGTATCTACTGG	chr12:114,512,371	-
OFF_B_7	3	GaCAGATCCcAGGTACTCACGGG	chr17:39,201,323	<i>RPL19</i>
OFF_B_8	3	GgCAGACCCAAGtTcCTCACTGG	chr17:49,244,982	-
OFF_B_9	3	cCCAGACCCagGGTcCTCACTGG	chr9:127,112,016	<i>RALGPS1</i>
OFF_B_10	3	GCCAAcTCAAGGTATCTACTGG	chrX:34,865,858	-
OFF_B_11	3	GCCAAcTCAAGGTtCTCACTGG	chr10:102,540,017	<i>SUFU</i>
OFF_B_12	3	GcTAGACaCAAGGTACTgACTGG	chr13:57,286,167	-
OFF_B_13	3	GCCAGACCCAAaGcACaCACAGG	chr19:2,775,589	<i>SGTA</i>
OFF_B_14	3	aCCAGACCCttGGTACTCACTGG	chr3:40,138,279	<i>MYRIP/EIF1B-AS1</i>

The presence of off-target genome editing events on the established cell line Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2 was also addressed in this thesis. First, the possible off-target positions (up to three mismatches) for the *BTBD8*_sgRNA1 were identified using the CasOFFinder⁶⁸ and Synthego's Verify Guide Designing webtools²⁸². According to both algorithms, the selected guide could potentially target 14 positions, recollected in Table 30. Once identified, primer pairs were designed to sequence those positions in the genome of both the edited cell line (Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2) and its isogenic WT control (Ctrl.BS-PBiPs37-Sv4F-1). The designed primer pairs can be found in Table 31.

Table 31. Sequencing primers designed to study the off-target regions of the *BTBD8*_sgRNA1.

TARGET	SEQUENCE (5->3)	PRIMER	ANNEALING TEMPERATURE	AMPLICON LENGTH
<i>TSHZ2</i>	TCAGGCGTGCAGTTTTTAATGC	OF_B_1_F	60	659
	CGTCCTAGAGGTCCATCCAAC	OF_B_1_R		
<i>LINC01716</i>	TCCTTCTGGGGTGTAGAGCAG	OF_B_2_F	60	941
	GTAATTAGCAAGGGCTAAACCAAC	OF_B_2_R		
-	ATACCCGCTCCACCCGTTT	OF_B_3_F	63	898
	CCAGTAGAGTGGGGCACTGCTG	OF_B_3_R		
-	ATACCCGCTCCACCCGTTT	OF_B_4_F	63	639
	CACAATTACCCCCGCCCC	OF_B_4_R		
<i>ENSG00000290113</i>	TCGCCGCTAAGAGTAAAAGTGGA	OF_B_5_F	60	1057
	TTCTTTCCGGCGACAATGA	OF_B_5_R		
-	ACTAAAAGATGGGGCCGGGGG	OF_B_6_F	63	1022
	GCCCTCTGAATCATGTGGGTG	OF_B_6_R		
<i>RPL19</i>	GGCATAGCGTGGGTTTTGAAG	OF_B_7_F	60	941
	AGCACACTCAAAGAACAAGCC	OF_B_7_R		
-	GGACCCGACTATTCATGTGCC	OF_B_8_F	63	874
	CCCAGTTTCTGCAAGCAGGGT	OF_B_8_R		
<i>RALGPS1</i>	CTGGTCTGCTGGATTGGAAC	OF_B_9_F	60	375
	CAAACCAGAAGCTCCACAC	OF_B_9_R		
-	AGGATGTGAAAGGCATCGGCA	OF_B_10_F	63	1150
	ACATGGGCAAAGGGACTTTGC	OF_B_10_R		
<i>SUFU</i>	TGTATTAGGGCTCTCTAGAAAAAC	OF_B_11_F	60	963
	GGAATGAATGCCACGGGTATAG	OF_B_11_R		
-	GAAAGCAACCCAAGCCCGTT	OF_B_12_F	60	793
	CGTGCAGGATCCACTAGG	OF_B_12_R		
<i>SGTA</i>	TGACCTGCCCCCACTCAA	OF_B_13_F	63	680
	CCGTAACCTCCTGTCCAGCG	OF_B_13_R		
<i>MYRIP/EIF1B-AS1</i>	GCAAGTGAGTGAACAATGAC	OF_B_14_F	55	833
	ATTAACCTAAGGGGAAACATTAAG	OF_B_14_R		

As can be observed in Figure 61, no alterations were found in 11 of the 14 potential off-target positions analyzed. Unfortunately, due to the presence of repetitive sequences, Sanger sequencing failed to provide any information regarding unexpected editing events in the off-target regions associated with OFF_B_6, OFF_B_8, and OFF_B_14 (Table 30).

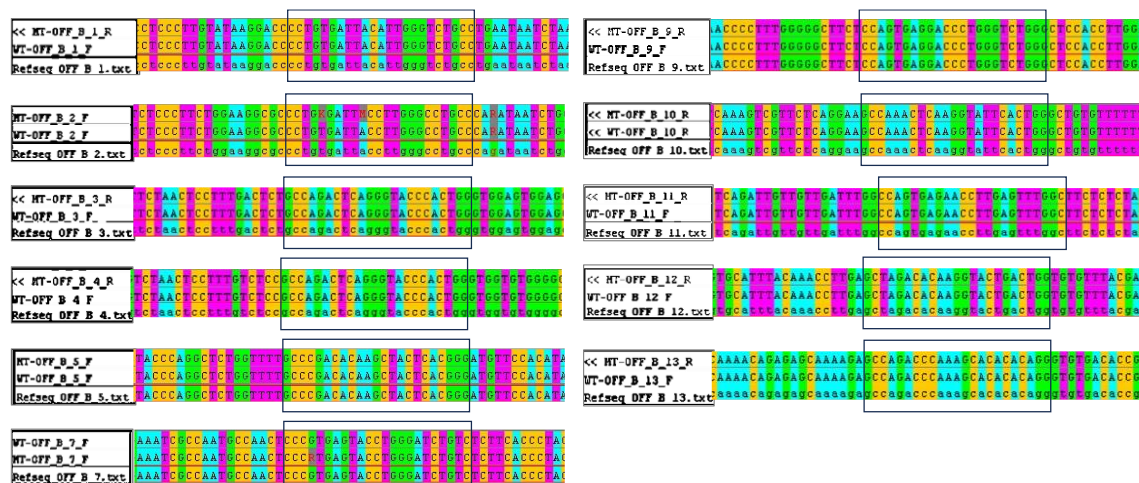


Figure 61. Analysis of the presence of editing events in off-target positions for the guide *BTBD8*_sgRNA1. The region of each potential off-target was sequenced in the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8*-i2 (indicated as “MT” in the image), and on the isogenic control Ctrl.BS-PBiPs37-Sv4F-1 (indicated as “WT” in the image). Both sequences were aligned with the genomic reference sequence, previously indicated in Table 30. Black rectangles indicate the predicted off-target region. Image created by the author in Codon Code Aligner v10.0.2.

4.1.3. Evaluation of the functional consequences of *BTBD8* mutations in human brain organoids.

In recent years, the development of human brain organoids has provided a valuable tool for studying human neurodevelopment and neurobiology. The main objective of this section is to assess the impact of *BTBD8* mutations on the normal development of these three-dimensional, *in vitro* models of the human brain. Identifying disruptions in the organoids' growth, structure, or function that could be linked to *BTBD8* mutations could be key to gain insights into the specific function of the gene and its role in normal neurodevelopment, as well as its association to neurological disorders.

4.1.3.1. Assessment of alterations in brain organoid size.

Photographs were taken with an optical microscope on medium-change days for the whole duration of the organoid differentiation protocol. By doing this, it was intended to evaluate the growth progression of the brain organoids derived from the mutant cell line Ctrl.BS-PBiPS37-SvF4-1-*BTBD8*-i2 in comparison to their WT isogenic control.

During the first 18 days of the protocol, it was possible to precisely follow each spheroid individually as they were separated in a 96-well plate. Representative images of the development of the *BTBD8*-mutant spheroids and their WT isogenic control can be found in Figure 63. From day 18 on, representative images of the organoids growing of 6-well plates were taken to continue following their development until they were too big to be fully captured with the smaller objective of the microscope (Figure 64). In those images, it is noticeable that *BTBD8*-mutant human brain organoids appear to be larger than the WT. As shown in Figure 62, in addition to the images, the area of the organoids was measured to be able to quantify the observed differences, which were significant from day 6 until day 35. From day 35 on, it was not possible to quantify the size differences since the *BTBD8*-mutant organoids were too large to be precisely measured.

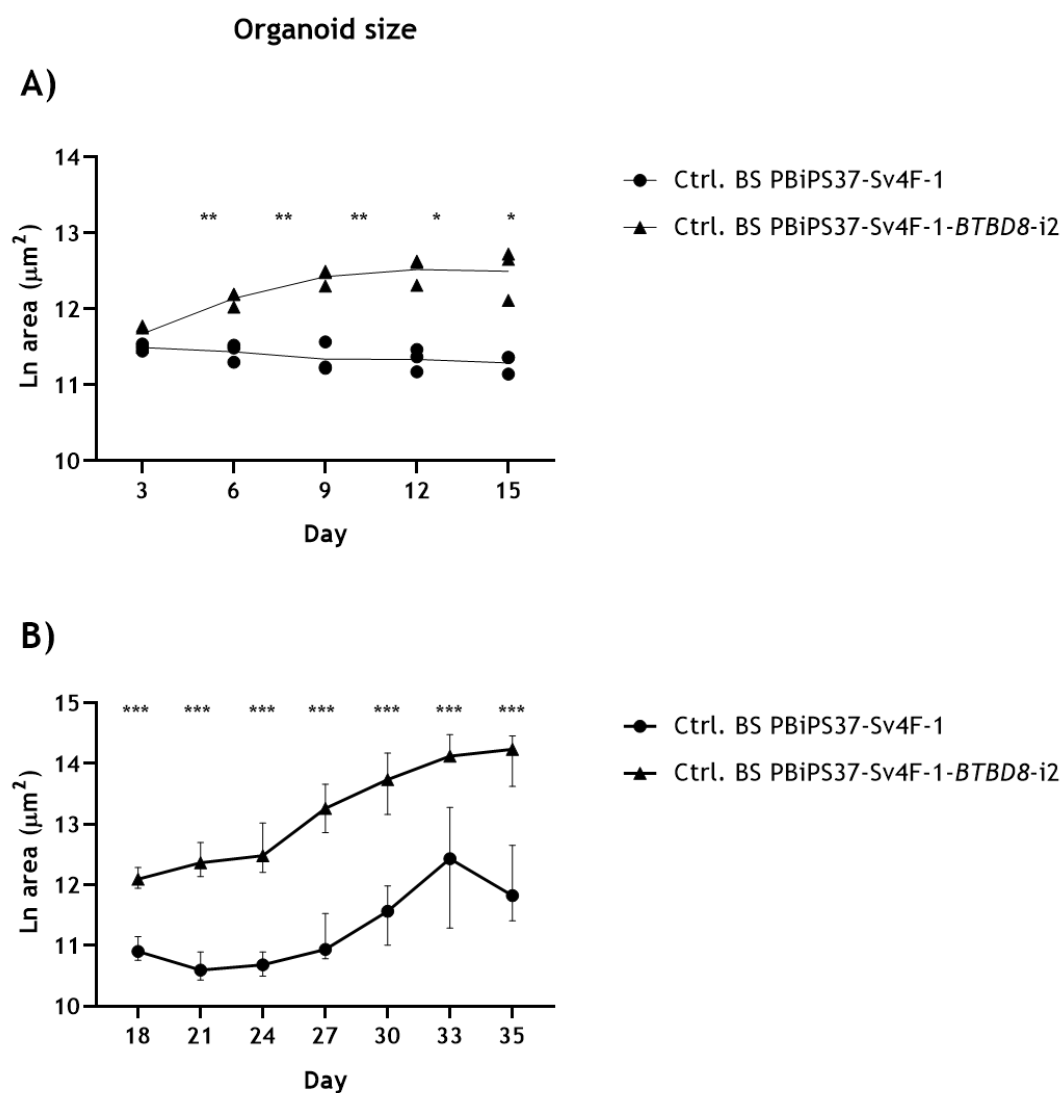


Figure 62. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and *BTBD8*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8-i2*) hiPSCs lines. A) Days 3-15 of the protocol.

B) Days 18-35 of the protocol. Area is represented in logarithm scale as data show a lognormal distribution. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

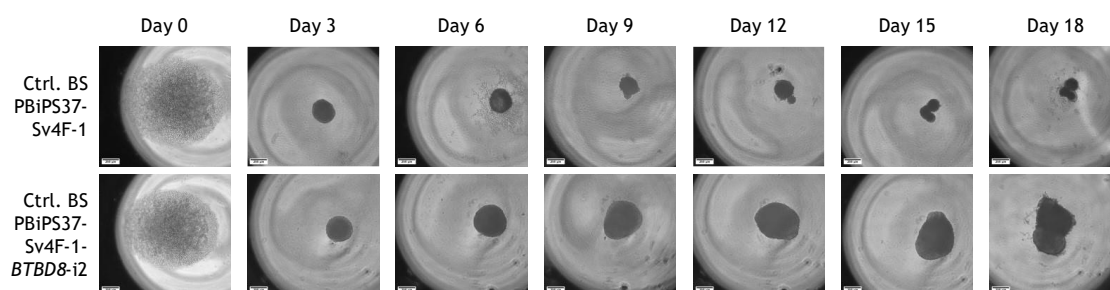


Figure 63. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and *BTBD8*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8-i2*) hiPSCs lines on the first 18 days of the protocol. The same well was imaged per cell line on each time point.

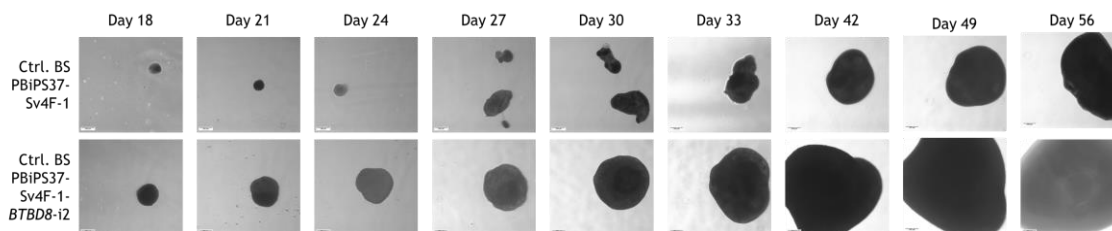


Figure 64. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and *BTBD8*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8-i2*) hiPSCs lines from day 18 to 56 of the protocol.

4.1.3.2. Immunohistochemistry characterization of human brain organoids harbouring mutations in *BTBD8*.

The internal structure and cell composition of human brain organoids derived from the cell line Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8-i2* was also addressed in this thesis.

First, the expression of the *BTBD8* protein was studied in the mutant brain organoids and their isogenic controls at three time points (day 45, 60 and 90) using an anti-KIAA1107 antibody. As can be observed in Figure 65, the protein could be detected in both the control and mutant organoids, although the expression is low. This indicates that the epitope is present, which was expected since one *BTBD8*-allele of the cell line Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8-i2* contains a stop gain mutation, but the other one has a small deletion (1 aa). Considering the presence of significant size differences with isogenic control organoids, two possible hypotheses were considered. The first hypothesis was that *BTBD8* was not functional, since the 1 aa deletion was predicted to alter its 3D structure, and thus could have important implications on the protein interactions and functionality. The second hypothesis was that phenotypic alterations could be caused in heterozygosis (only one non-functional copy of *BTBD8*), which was supported by the observations of Alonso et al^{26,216}.

To further investigate the presence of structural or molecular alterations that could support the hypothesis of an altered *BTBD8* function in the mutant organoids, slices were stained against key markers of neuron development. No differences were found between the two models for the dorsal forebrain neural progenitor marker PAX6, or the mature pan-neuronal marker MAP2 (Figure 66 and Figure 68). Next, it was analysed the presence of rosette-like structures enriched in neural progenitors (SOX2⁺ PAX6⁺) in both the WT and mutant models, which are expected to decrease as the organoids mature. At the structural level, it was noticeable the presence these structures in both, although in the *BTBD8*-mutant organoids they are more peripheral and less abundant. The expression of SOX2 was also quantified in the models, revealing that the number of SOX2⁺ cells at day 45 in *BTBD8*-mutant organoids is more similar to the one observed at day 62 in the WT control. Taken together with the absence of differences at day 60, this indicates that the mutant model neurodevelopment might be accelerated in comparison to the WT.

Alterations in the formation of GABAergic neurons were also identified in the *BTBD8*-mutant model, as day 45 and 60 organoids show less expression of DLX2⁺ cells, a marker of GABAergic interneuron precursors. The expression of the GABAergic marker GAD67 was expected to increase from day 45 to day 60 as the organoids mature, which was the case for the WT model. However, *BTBD8*-mutant organoids maintain the same level of expression between the two time points, also suggesting alterations in the GABAergic neurogenesis (Figure 67, Figure 68).

Finally, the expression of two subcortical projection neuron markers was studied in the obtained human brain organoids. Deep-layer (CTIP2/BCL11B⁺) and upper-layer (SATB2⁺)

neurons were identified, with no apparent differences between the BTBD8-mutant and the WT isogenic control (Figure 69).

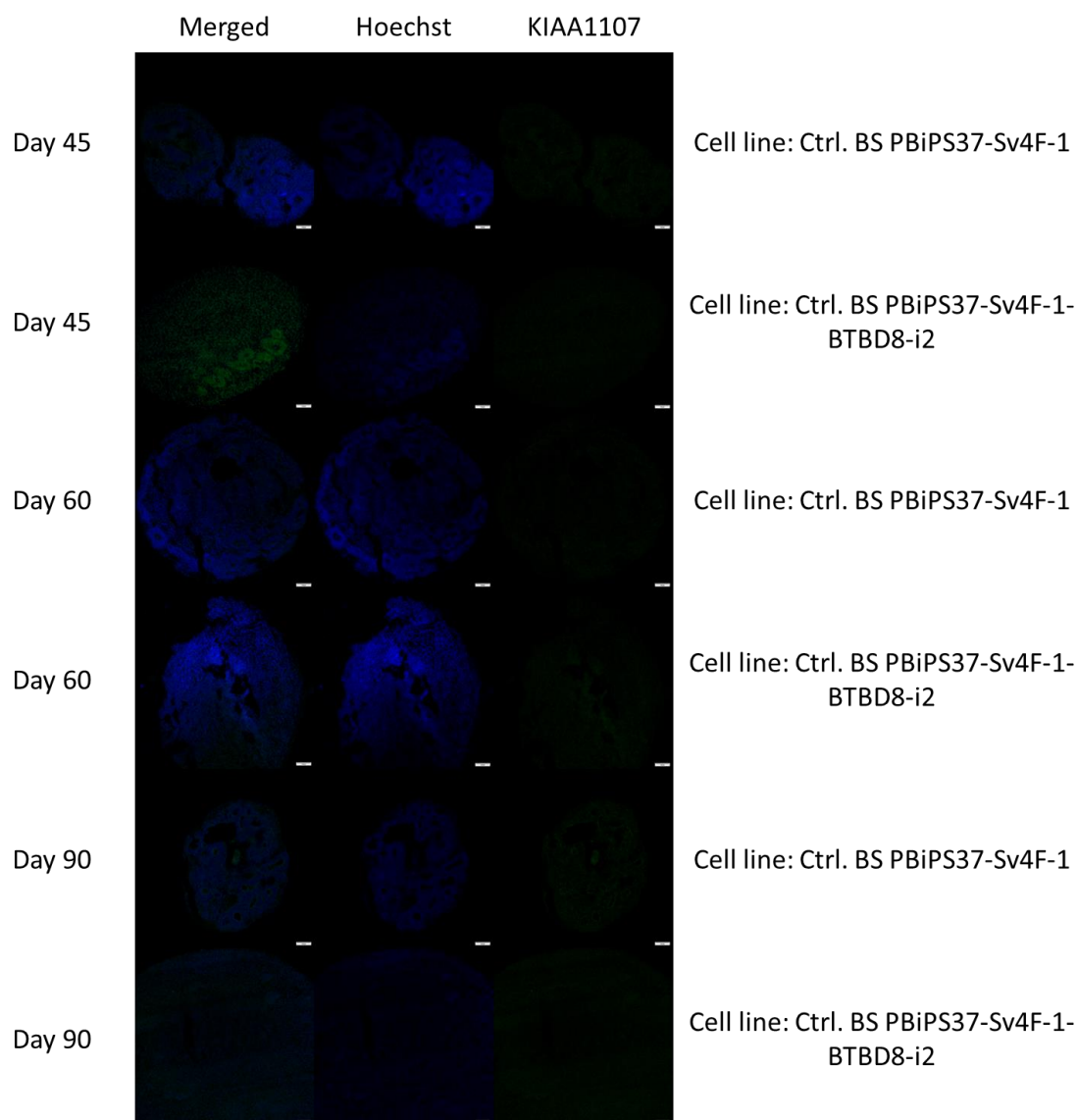


Figure 65. Immunostaining of the antigen KIAA1107 in brain organoids derived from the edited cell line Ctrl.BS-PBiPS37-SvF4-1-BTBD8-i2 and their isogenic wild type control (Ctrl.BS-PBiPS37-SvF4-1). Nuclear staining was performed with Hoechst.

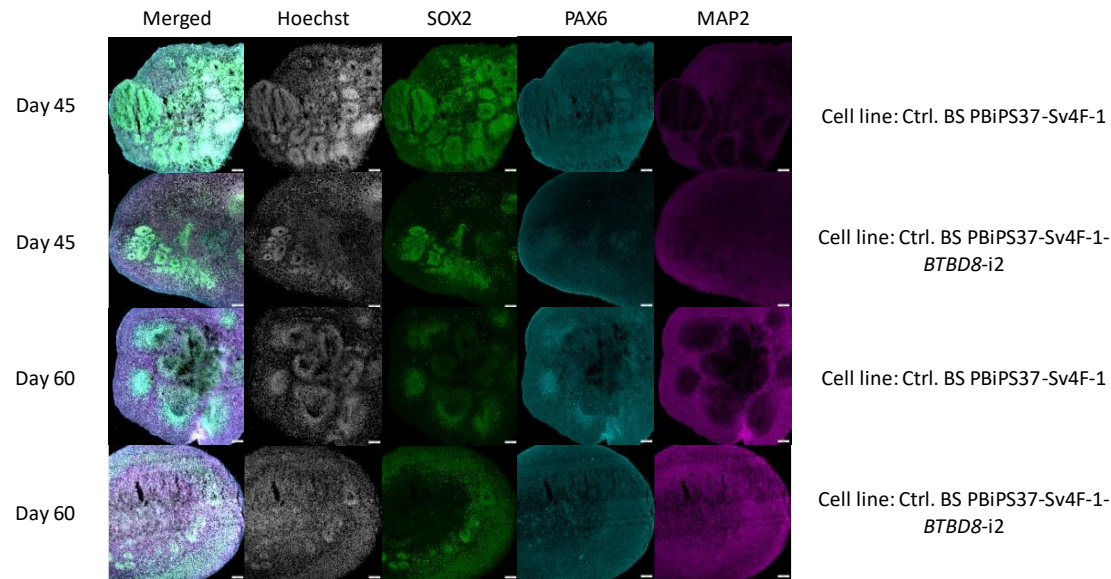


Figure 66. Characterization of the expression of cell-markers SOX2, PAX6, and MAP2 by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and *BTBD8*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8*-i2) cell lines. Nuclear staining was performed with Hoechst.

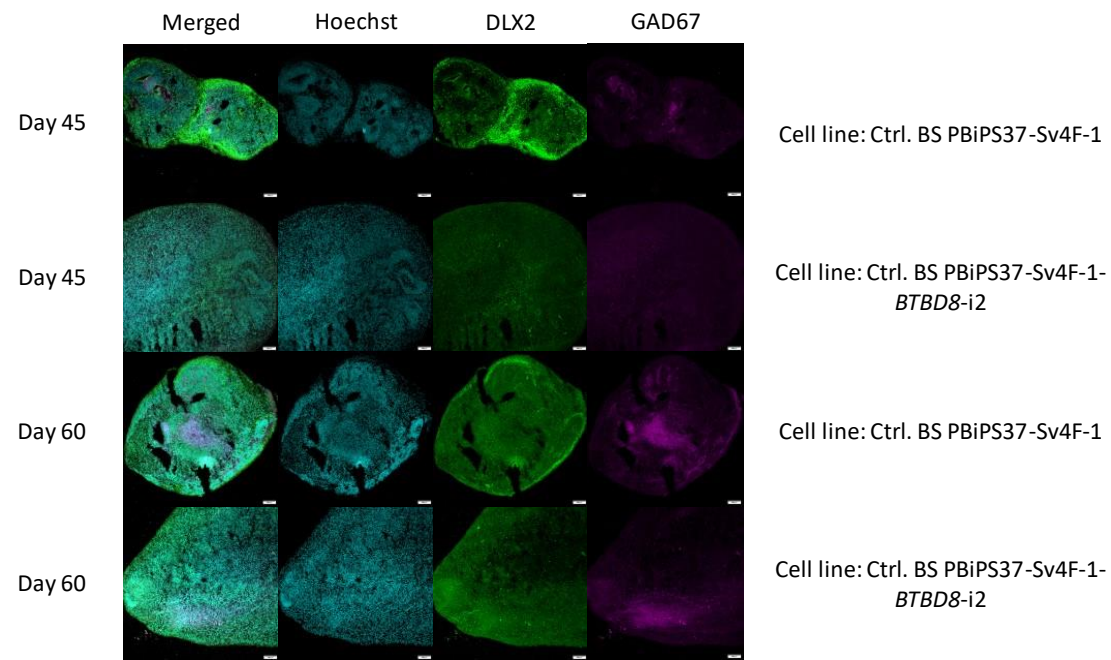


Figure 67. Characterization of the expression of cell-markers DLX2 and GAD67 by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8*-i2) cell lines. Nuclear staining was performed with Hoechst.

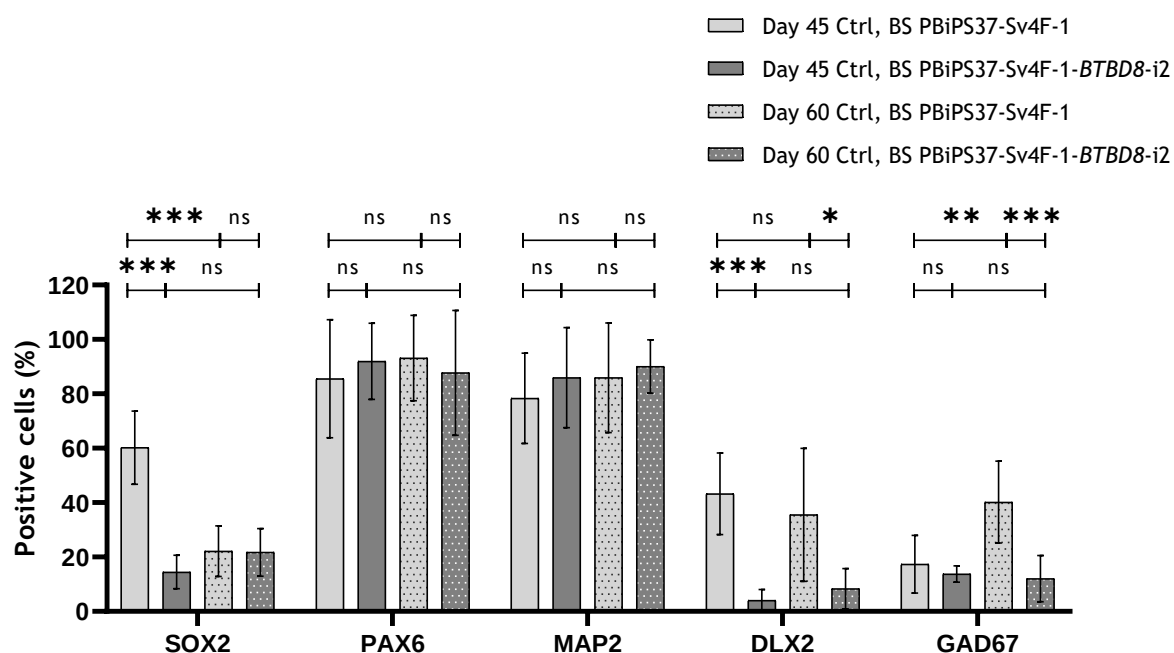


Figure 68. Quantification of the expression the cell-markers SOX2, PAX6, MAP2, DLX2, and GAD67 in human brain organoids. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and *BTBD8*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8*-i2) cell lines. Data represented as mean $\pm$ SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

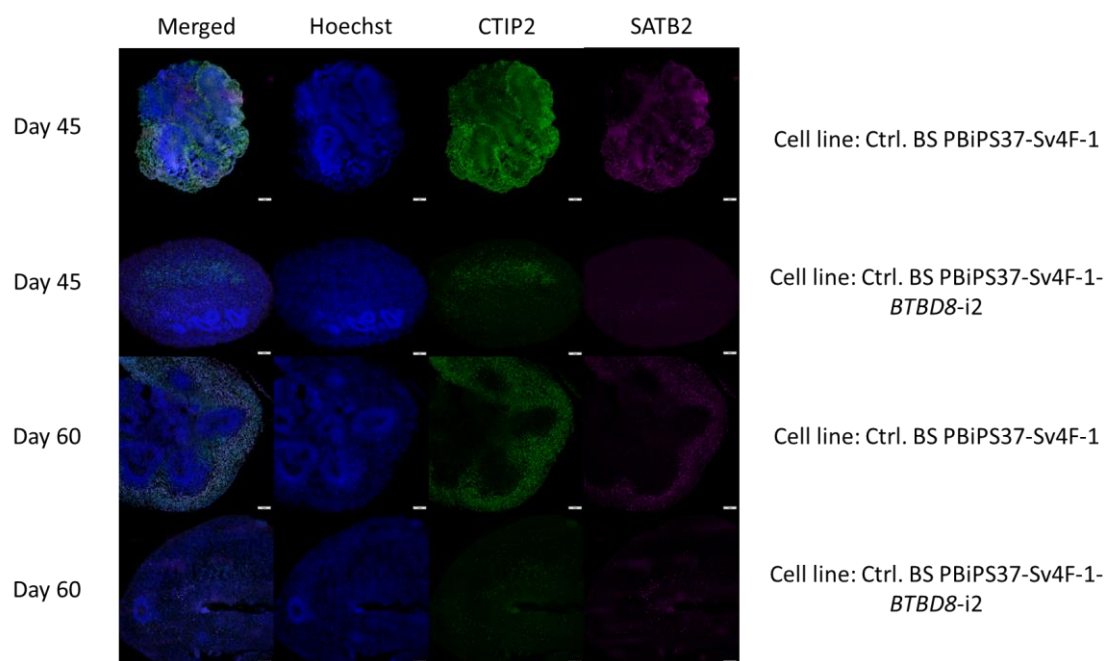


Figure 69. Characterization of the expression of cell-markers CTIP2 and SATB2 by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8*-i2) cell lines. Nuclear staining was performed with Hoechst.

4.1.4. Evaluation of neurodevelopmental alterations associated with mutations in *BTBD8* in cortical neurons derived from hiPSCs.

The second approach used in this thesis to study the impact of *BTBD8* mutations in brain development involved the generation of cortical neurons from the cell line Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8-i2*, previously generated and characterized. The scope of the present work included the optimization of the cortical differentiation protocol and the evaluation of morphological alterations in the obtained neurons. Finally, a proteomic approach was applied to determine whether mutations in *BTBD8* are sufficient to induce molecular changes in human neurons leading to neurodevelopmental alterations.

4.1.4.1. Characterization of cortical-like neurons derived from hiPSCs.

The characterization of the cell types obtained after hiPSCs differentiation is key to evaluate the success and suitability of the selected protocol. In this thesis, immunocytochemistry was applied for identifying the presence of cortical neurons in cultures differentiated from the WT cell line Ctrl.BS-PBiPS37-Sv4F-1.

In Figure 70 can be observed that the pan-neuronal marker MAP2 is expressed extensively throughout the differentiated cultures, supporting the neuronal-commitment of the obtained cells. In that same figure is also shown that most cells are positive for PAX6, a transcription factor that can be used as a marker for neuronal progenitor cells from the cortex, suggesting that the obtained neuronal cultures are still immature after 60 days of differentiation with the followed protocol. The expression of the progenitor marker SOX2 further supports this hypothesis (also in Figure 70).

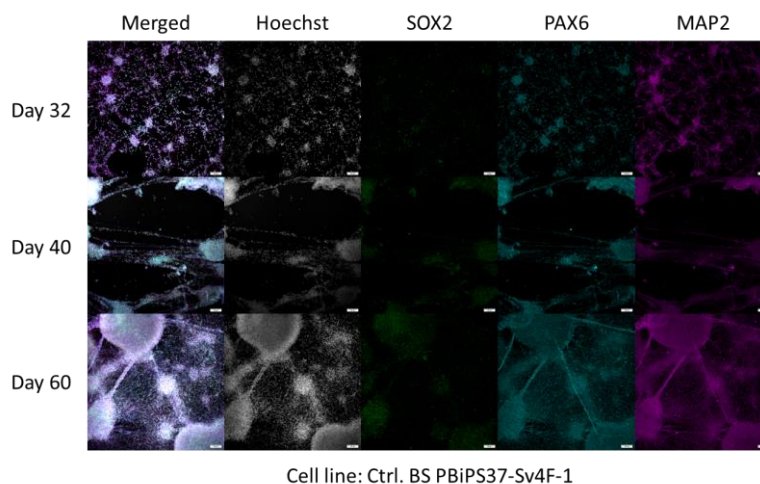


Figure 70. Characterization of the expression of cell-markers SOX2, PAX6, and MAP2 in cortical-like neurons derived from the Ctrl.BS-PBiPS37-Sv4F-1. Nuclear staining was performed with Hoechst.

Next, the expression of the subcortical projection neuronal markers CTIP2/BCL11B and SATB2 was assessed. As shown in Figure 71, 40 days after the beginning of the differentiation process, early-born deep layer CTIP2/BCL11B⁺ cells are abundant in the cultures, while later-born upper-layer SATB2⁺ are still rare. In Figure 72 can also be observed that most cells are positive for vGlut1, a specific protein of excitatory glutamatergic neurons.

Finally, the expression of *BTBD8* was verified in the differentiated neuronal cultures at different time points. By doing so, it was intended to investigate the suitability of this *in vitro*

model to further characterize the role of the protein in cortical development and its alterations. In Figure 72 can be observed that BTBD8 was detected in the obtained cortical populations, with increased expression early on the differentiation process.

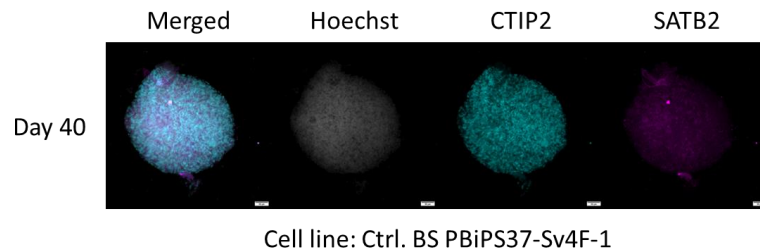


Figure 71. Characterization of the expression of cell-markers CTIP2/BCL11B and SATB2 in cortical-like neurons derived from the Ctrl.BS-PBiPS37-Sv4F-1. Nuclear staining was performed with Hoechst.

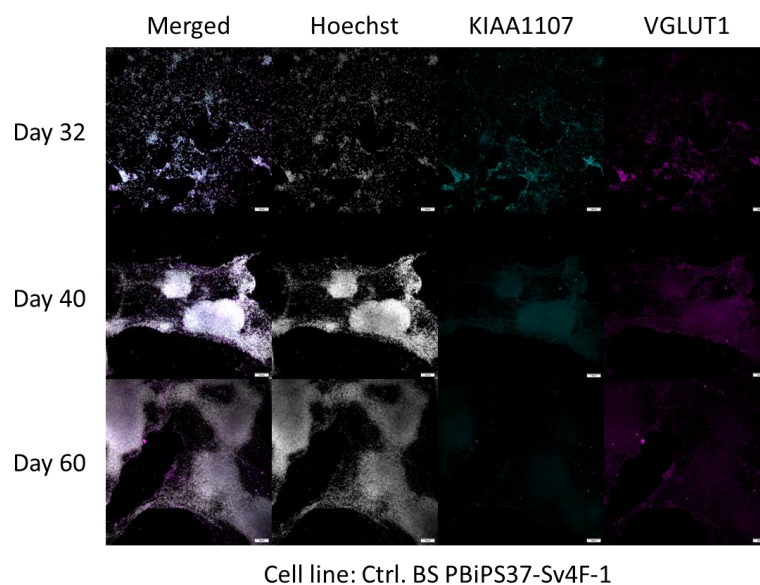


Figure 72. Characterization of the expression of cell-markers BTBD8 and VGLUT1 in cortical-like neurons derived from the Ctrl.BS-PBiPS37-Sv4F-1. Nuclear staining was performed with Hoechst.

4.1.4.2. Morphometric analysis of cortical neurons derived from *BTBD8*-mutant hiPSCs.

Once the cortical differentiation protocol had been optimized, it was applied to obtain cortical neurons from the mutant cell line Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8*-i2. During this process, it was remarkable that *BTBD8*-mutant cultures were less homogeneous than the WT controls (Figure 73), which diffculted the identification of cells with neuron-like morphology. To address this problem, the *BTBD8*-mutant neurons and their isogenic controls were transfected with a GFP-encoding plasmid to visualise and study the morphology of isolated cells from the neuron differentiating cultures. By doing so, it was possible to identify cells with neuronal morphology in cultures derived from both cell lines. As indicated in Figure 74A, no significant differences were found between *BTBD8*-mutant neurons and the wild type controls the soma size after 43 or 63 days of differentiation. In addition, no alterations were detected in the dendritic branching, neither on the total number of dendrites on each neuron (Figure 74B),

or in their complexity (Figure 74C) at the same time points. Representative images of the analysed GFP-transfected neurons can be found in Figure 75.

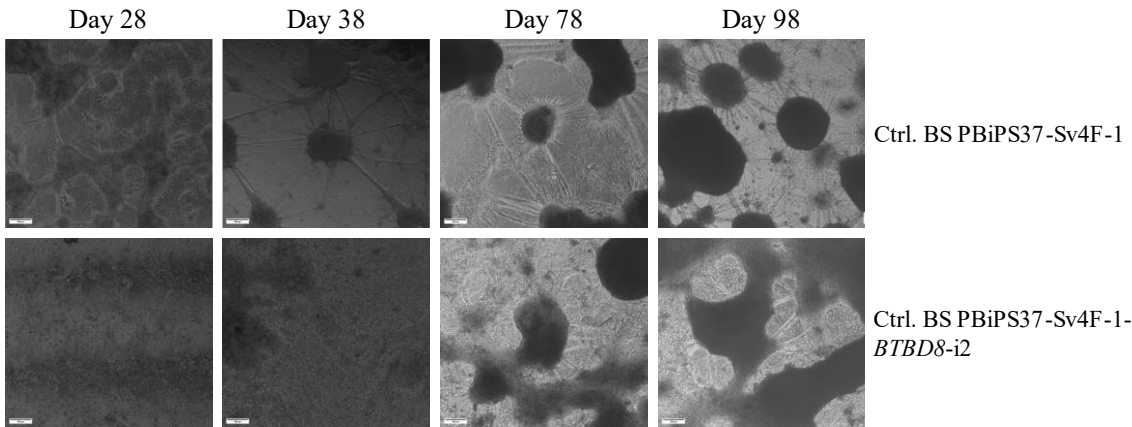


Figure 73. Differentiation of cortical neurons from the cell lines Ctrl.BS-PBiPS37-Sv4F-1 and Ctrl.BS-PBiPS37-Sv4F-1-BTBD8-i2. Nuclear staining was performed with Hoechst.

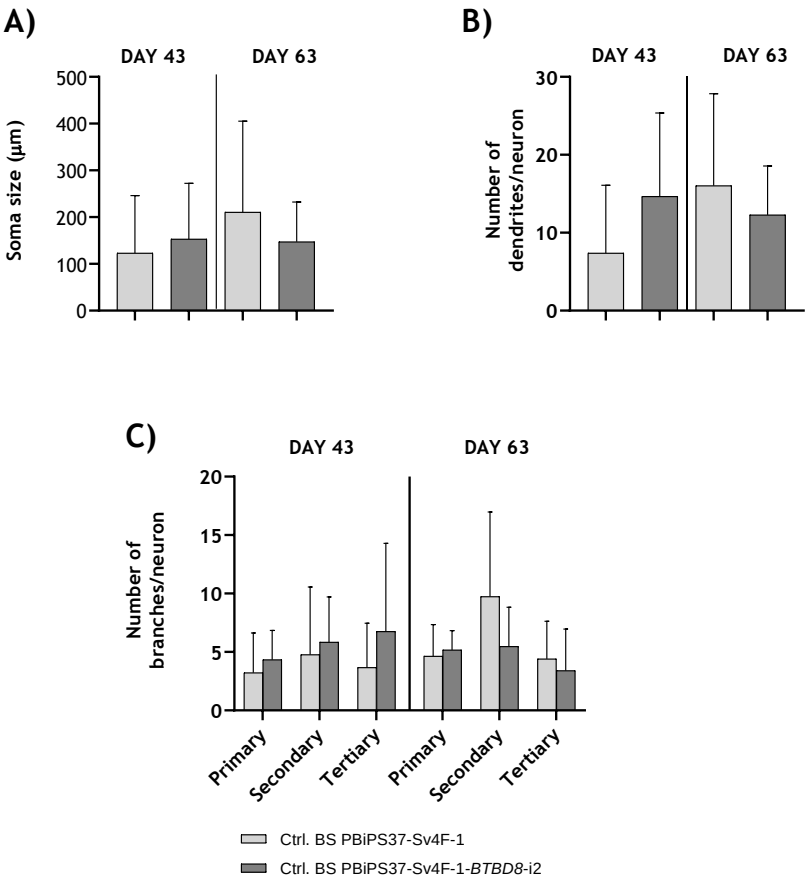


Figure 74. Morphometric analysis of neurons in monolayer derived from the cell line Ctrl.BS-PBiPS37-Sv4F-1-BTBD8-i2 and its isogenic control Ctrl.BS-PBiPS37-Sv4F-1. A) Comparison of soma sizes. B) Comparison of the number of dendrites on each neuron. C) Comparison of the number of branches on each neuron. Data represented as mean + SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

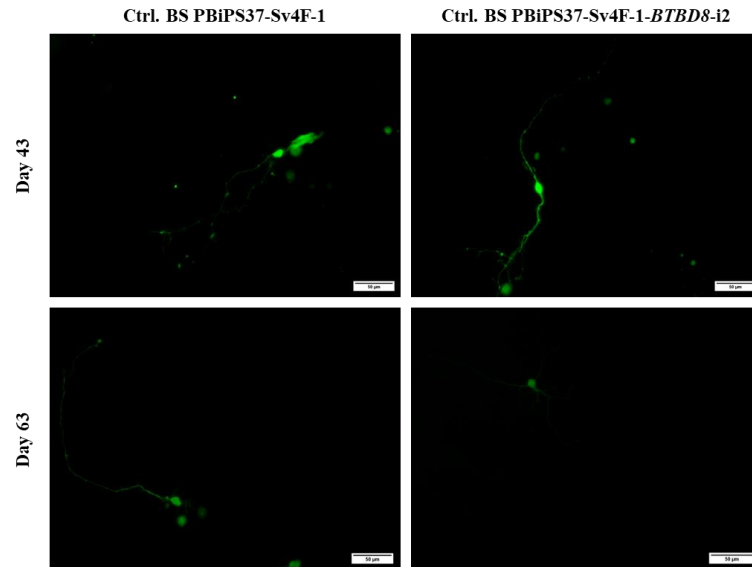


Figure 75. Representative images of GFP-transfected neurons in monolayer.

4.1.4.3. Proteomic analysis of cortical neurons derived from *BTBD8*-mutant hiPSCs.

Samples obtained from cortical-like neurons, cultured for 40 and 60 days, were used to identify proteomic differences between the *BTBD8*-mutant line (Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8-i2*) and its isogenic control (Ctrl.BS-PBiPS37-Sv4F-1). The Venn diagram represented in Figure 76 shows the number of proteins identified on each sample with a false discovery rate (FDR) of 1 %. Overlapping regions in the Venn diagram indicate the number of shared proteins between samples.

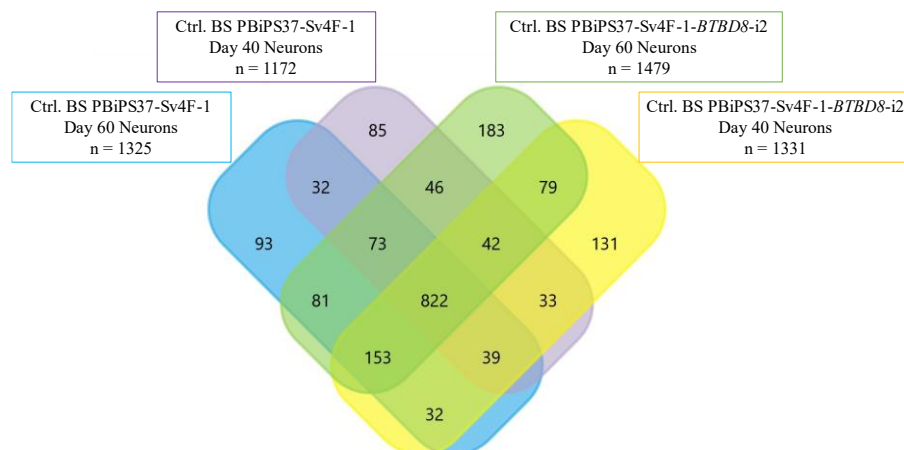


Figure 76. Venn diagram of the number of proteins detected on each sample. False discovery rate 1 %.

The proteomic data was explored to identify proteins of interest, starting with *BTBD8* (Figure 77). This protein was detected in wild-type neuronal cultures differentiated for 40 days, but not in later time points (day 60). In comparison, *BTBD8*-mutant cultures showed expression of *BTBD8* at both time points, with its levels increasing over time. One possible explanation to those observations was an impaired degradation of *BTBD8* that explained its accumulation over time. Another possible explanation could be a delayed cortical differentiation of *BTBD8*-mutant hiPSCs, thus maintaining *BTBD8*'s expression longer.

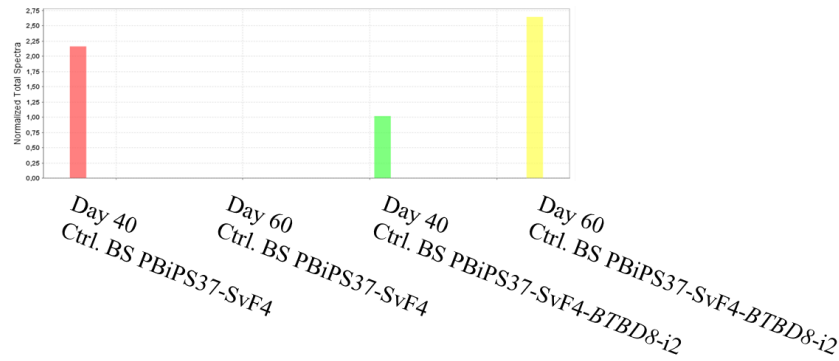


Figure 77. Expression levels of BTBD8 in the proteomic assay. Normalization data was obtained taking Day 40 Ctrl.BS-PBiPS37-SvF4 as the reference sample. Quantification was performed in Scaffold v5.3.1. Red: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Blue: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Green: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4-BTBD8-i2; Yellow: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4-BTBD8-i2.

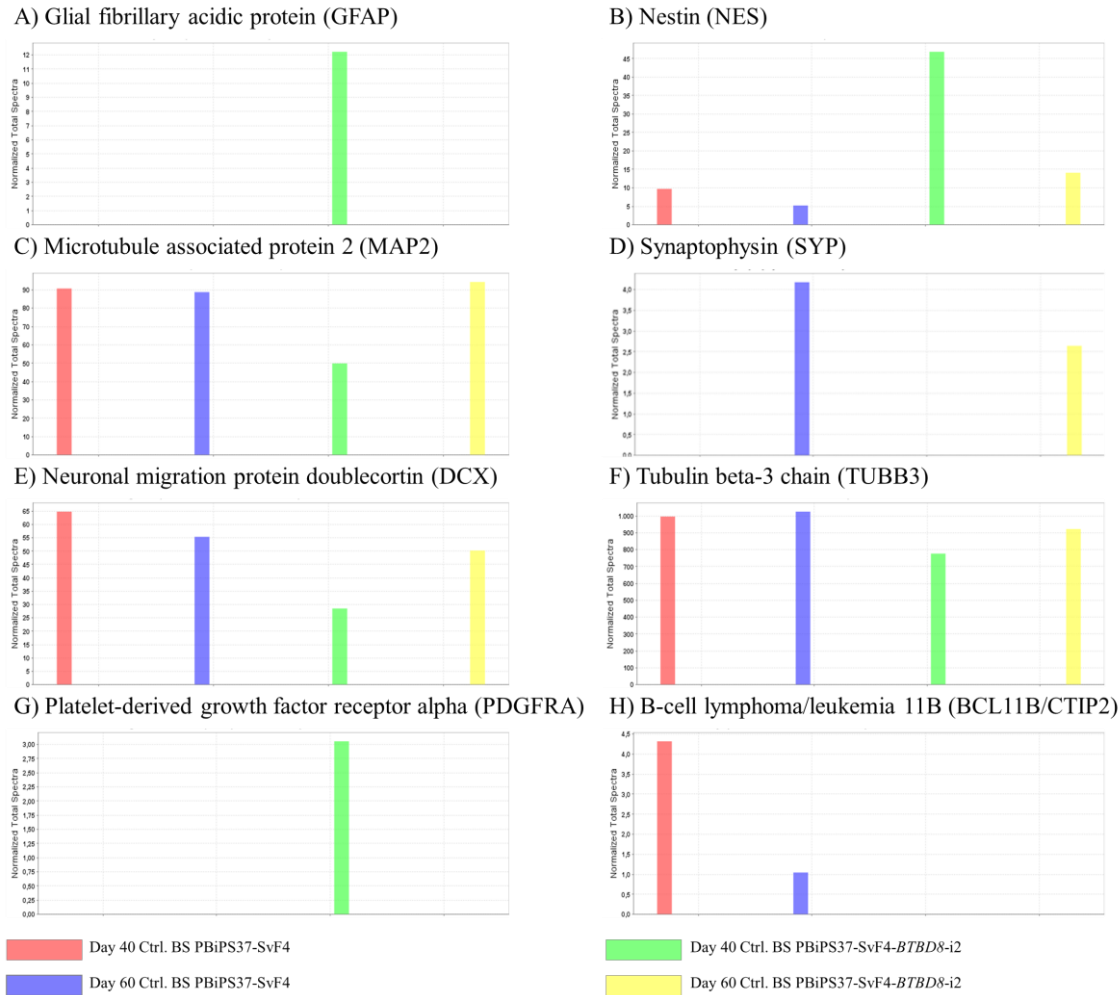


Figure 78. Expression levels of cell-type markers of interest in the proteomic assay. Normalization data was obtained taking Day 40 Ctrl.BS-PBiPS37-SvF4 as the reference sample. Quantification was performed in Scaffold v5.3.1. Red: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Blue: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Green: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4-BTBD8-i2; Yellow: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4-BTBD8-i2.

Next, neuronal cell-type markers' expression was interrogated in the proteomic data to investigate the possibility of a delayed cortical differentiation (Figure 78). Neural stem cell markers (GFAP, NES) were almost absent in wild type cultures, both at day 40 and 60. The same pattern was observed for the platelet-derived growth factor receptor alpha, an oligodendrocyte precursor marker. In comparison, those markers were strongly detected in *BTBD8*-mutant cultures differentiated for 40 days (Figure 78 A, B, G).

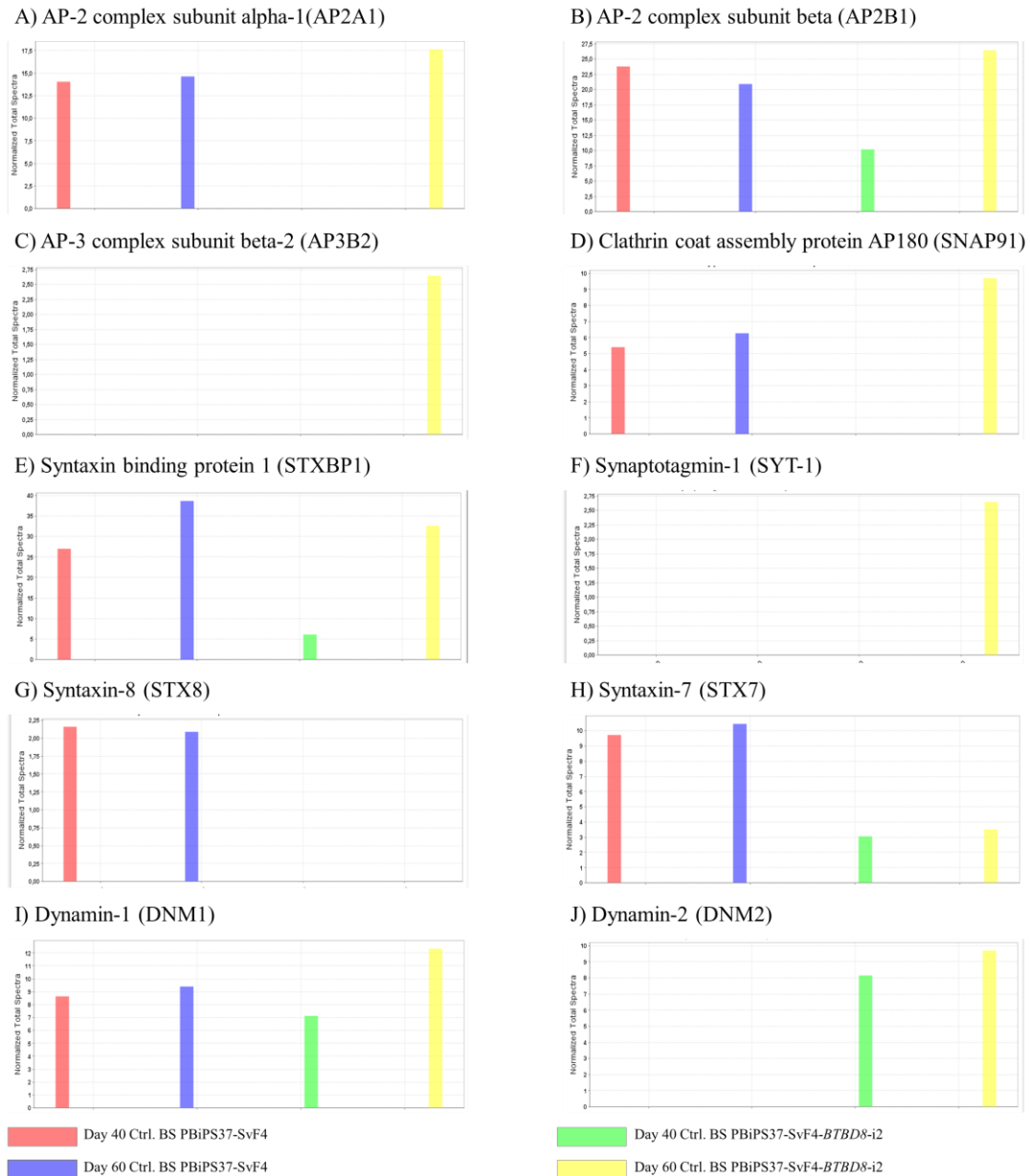


Figure 79. Expression levels of *BTBD8*-associated proteins implicated in endo/exocytosis pathways. Normalization data was obtained taking Day 40 Ctrl.BS-PBiPS37-SvF4 as the reference sample. Quantification was performed in Scaffold v5.3.1. Red: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Blue: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Green: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4-*BTBD8*-i2; Yellow: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4-*BTBD8*-i2.

Immature neuronal markers (DCX, TUBB3) were detected at both time points of interest in wild type and *BTBD8*-mutant cultures, although the expression seems reduced in the latter (Figure 78 E, F). This was also observed for the pan-neuronal marker MAP2 (Figure 78 C). The

mature-neuronal marker SYP was only detected after 60 days of differentiation in both cell lines, consistent with an increased maturation of the cultures. It is worth noticing that despite being detected, SYP levels appear to be reduced in *BTBD8*-mutant cultures (Figure 78 D).

Finally, CTIP2/BCL11B was more abundant at day 40 than at day 60 in wild type cultures, consistent with the formation of early-born, deep layer subcortical projection neurons. The expression of this marker could not be detected in *BTBD8*-mutant cultures (Figure 78 H).

Altogether, these results indicate that cultures differentiated from the cell line Ctrl.BS-PBiPS37-Sv4F-1-*BTBD8-i2* show an enrichment in progenitor cells and immature neurons, with apparent impairments in the formation of subcortical projection neurons.

The proteomic analysis continued with the characterization of predicted *BTBD8*-interactors, based on information gathered from publications and public databases^{211,290,299}. First, proteins implicated in exocytosis and endocytosis pathways were studied, since the role of *BTBD8* had been previously associated with synaptic vesicle trafficking (see section 1.8.1.1). In Figure 79 can be observed the expression of the potential interactors that participate in those pathways: AP2A1, AP2B1, AP3B2, SNAP91, STXBP1, SYT1, STX8, STX7, DNM1, and DNM2.

BTBD8-mutant cultures differentiated for 40 days did not show expression of AP2A1 or SNAP91 (Figure 79A, D), two components of the adaptor complex that participate in clathrin-coated vesicles' endocytosis. Similarly, day 40 *BTBD8*-mutant cultures showed reduced expression levels of AP2B1 (Figure 79B), another adaptor complex subunit needed in that process. This pattern was also observed for STXBP1 (Figure 79E), a protein implicated in synaptic vesicle docking and fusion to the plasma membrane, and DNM1 (Figure 79I), a GTPase involved in vesicle trafficking and receptor mediated endocytosis.

Interestingly, the GTPase DNM2, likely involved in vesicle trafficking and receptor mediated endocytosis, was only detected in *BTBD8*-mutant cultures (Figure 79J).

AP3B2 and SYT1 were also detected only in *BTBD8*-mutant cultures, but just after 60 days of differentiation (Figure 79C, F). AP3B2 is part of an adaptor complex involved in intracellular vesicle trafficking, while SYT1 acts as a calcium sensor for neurotransmitter release.

Finally, alterations in two syntaxins involved in vesicle trafficking were identified: STX7 levels were reduced in *BTBD8*-mutant cultures, while STX8 was absent for them (Figure 79G, H).

After evaluating alterations in proteins involved in vesicle trafficking, the focus was shifted towards other potential interactors of *BTBD8*: DCTN1, ATP6V1A, CUL3, BCLAF1, NRP1, NRP2, AHNAK, and MAPT (Figure 80).

Dynactin subunit DCTN1 participates in microtubule trafficking and synaptic vesicle recycling. Despite being present in all samples, its levels were reduced in day 40 *BTBD8*-mutant neurons (Figure 80A). This pattern was also observed in MAPT (Figure 80H), a protein involved in microtubule assembly and neuronal polarity establishment.

The vacuolar ATPase component ATP6V1A was detected in all samples, with levels increasing over time in both cell lines. Nevertheless, a slight reduction was observed in day 60 *BTBD8*-mutants in comparison to the wild type control (Figure 80B).

CUL3 (Figure 80C), part of the cullin-RING-based BCR E3 ubiquitin ligase complex, was increased in *BTBD8*-mutant cultures. In contrast, *BTBD8*-mutant cultures had AHNAK depletion, a protein previously linked to neuronal cell differentiation (Figure 80G). Day 40 *BTBD8*-mutant neurons were also depleted in BCLAF1, a transcriptional repressor involved in apoptosis (Figure 80D).

Finally, cell-surface receptors NRP1 and NRP2 were only identified in *BTBD8*-mutant cultures. Although present at day 40, NRP1 levels were mainly detected in day 60 cultures, while NRP2 was only detected at the latest time point (Figure 80G, H).

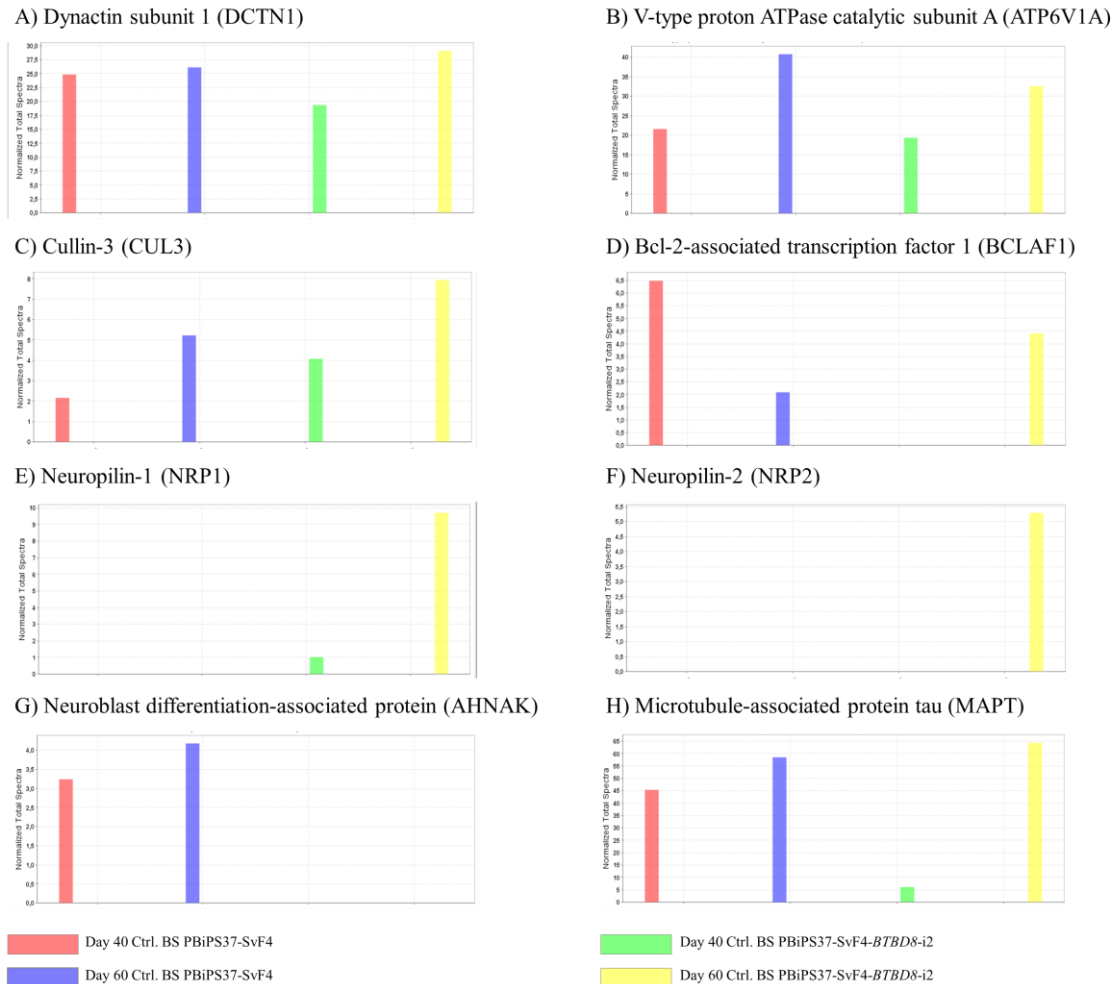


Figure 80. Expression levels of *BTBD8*-associated proteins of interest. Normalization data was obtained taking Day 40 Ctrl.BS-PBiPS37-SvF4 as the reference sample. Quantification was performed in Scaffold v5.3.1. Red: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Blue: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Green: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4-*BTBD8-i2*; Yellow: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4-*BTBD8-i2*.

The final part of the proteomic analysis was the conduction of pairwise comparisons between the proteins expressed by WT and *BTBD8*-mutant neurons on each time point (Annexe VI) with the purpose of identifying the biological meaning of those differences. These analyses revealed multiple alterations in *BTBD8*-mutant neurons, at many levels. In terms of morphological alterations, *BTBD8*-mutant neurons were depleted in dendrite associated proteins, while also being enriched in axon initial segment related proteins. In addition, *BTBD8*-mutant neurons were depleted in proteins associated with clathrin mediated endocytosis and phosphatidylinositol biosynthesis. Also noticeable were alterations in neurotransmission, such as impaired glutamate clearance and enrichment in proteins associated with GABA receptors, and cell signalling pathways, involving impairments in Reelin and PI3K/Akt pathways. Finally, alterations in ubiquitin-mediated protein degradation, purine catabolism, and gene expression were also identified.

Overall, these results indicate that *BTBD8* mutations (NC_000001.11:g.92177279_92177288delCTCACAGGAA and NC_000001.11:g.92177281_92177283delCAC) lead to an altered proteomic profile in cortical neurons derived from hiPSCs in monolayer. The relationship between the proteomic alterations mentioned in this section, *BTBD8* levels, and ASD development will be explored in the discussion.

4.2. GENERATION AND CHARACTERIZATION OF AN *IN VIVO* MODEL TO STUDY THE FUNCTION OF *FOSL2*.

As mentioned earlier in the introduction, Cospain and collaborators²²¹ have recently reported an association between *FOSL2* mutations and the presence of a syndromic developmental disorder in humans. To address the feasibility of using zebrafish as an *in vivo* model to study this neurodevelopmental condition, literature on the field was reviewed.

To date, only two publications have been published linking *fosl2* and zebrafish. The first one, by Kubra *et al.*³⁰⁰, explores the presence of members of *FOS* orthologous genes in *Danio rerio*, while the other one investigates the role of *fosl2* in myocardial differentiation. On this second article, published by Jahangiri *et al.*²³⁸, the authors knocked down *fosl2* using morpholinos and reported the cardiac effects of this downregulation. Despite not being reported in the article, cephalic alterations were also apparent in the published bright-field images of the morphant individuals, suggesting that *fosl2* can also be involved in neurodevelopment in this animal model. To empirically verify this observation, a knockdown experiment was performed in the laboratory using a splice blocking morpholino (GeneTools) targeting the first splice donor site in the *fosl2* pre-mRNA.

4.2.1. Mortality after morpholine treatment.

At 48 h post morpholine injection, the number of surviving embryos was quantified (Table 32). As a result, it was noticeable that the mortality rate in the morphant groups was much higher than the observed in the control. This mortality can be in part explain due to p53-mediated neurotoxicity. However, considering that the highest mortality was detected in the group only treated with MO-*fosl2*, this observation might be mainly associated with the knockdown of *fosl2*.

Table 32. Mortality rate 48 h after morpholine treatment in zebrafish embryos.

EXPERIMENTAL GROUP	NUMBER OF EMBRYOS (0 HPF)	NUMBER OF EMBRYOS (48 HPF)	MORTALITY (%)
Control	70	66	5,71 %
MO- <i>fosl2</i>	160	28	82,5 %
MO- <i>fosl2</i> + MO-p53	170	77	54.71 %

4.2.2. Knockdown verification in *fosl2* morphants.

On this knockdown experiment, a splice blocking morpholino was used to knockdown the expression of *fosl2*. More specifically, this morpholino targets the exon 1/intron 1 splice junction, blocking the action of the spliceosome in that region. Therefore, intron 1 (3628 bp) cannot be removed from *fosl2* pre-mRNA (Figure 81).

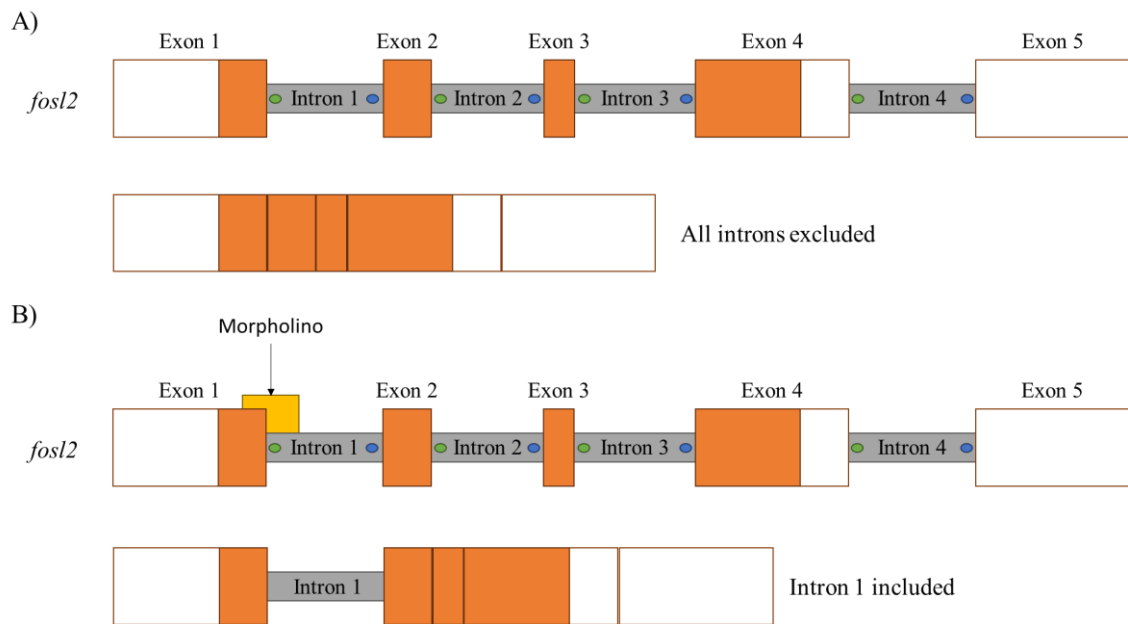


Figure 81. Expected outcome of the knockdown with the splice blocking morpholino that targets the first splice donor site. A) Splicing of the *fosl2* pre-mRNA under normal circumstances. B) Inclusion of intron 1 in the *fosl2* pre-mRNA due to the action of the splice blocking morpholino. Coding regions within exons are represented in orange. The target region for the morpholino is represented in yellow. Green and blue dots indicate the positions on pre-mRNA of the splice junctions, which are recognized by the spliceosome for correct processing.

The presence of altered *fosl2* pre-mRNAs was verified by conducting a RT-PCR using cDNA from control or *fosl2* morphant individuals as a template. Primers were designed to bind to exon 1 and exon 2, so the presence of intron 1 could be easily detected due to the size increase in the obtained amplicons. If control *fosl2* cDNA was amplified, a 193 bp amplicon should be obtained, while in morphants should be detected an amplicon of 3821 bp due to the inclusion of intron 1. The obtained amplicons were visualized on an agarose gel (Figure 82), on which it was confirmed that *fosl2* morpholino was indeed acting as expected, thus resulting in a non-functional protein.

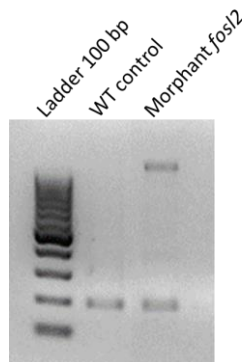


Figure 82. Agarose gel (1.5 %) loaded with the amplicons obtained after RT-PCR of cDNA samples from wild type control zebrafish and *fosl2* morphant individuals.

4.2.3. Phenotypic characterization of *fosl2* morphant embryos.

4.2.3.1. Morphometric analysis at 48 hpf.

The surviving embryos at 48 hpf (48 h after morpholine injection) were photographed, and morphometric parameters of neurodevelopmental interest were measured. Here, the full body length of the individuals was analysed, as well as the area and diameter of the cephalic and ocular regions, to identify alterations in the normal neurodevelopment of the zebrafish embryos. In Figure 83, representative photos taken of embryos from the three experimental conditions (control, MO-*fosl2*, MO-*fosl2* + MO-*p53*) are displayed.

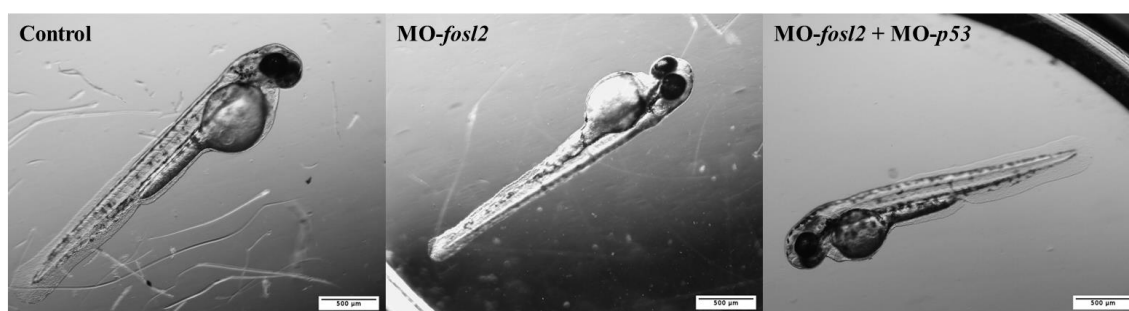
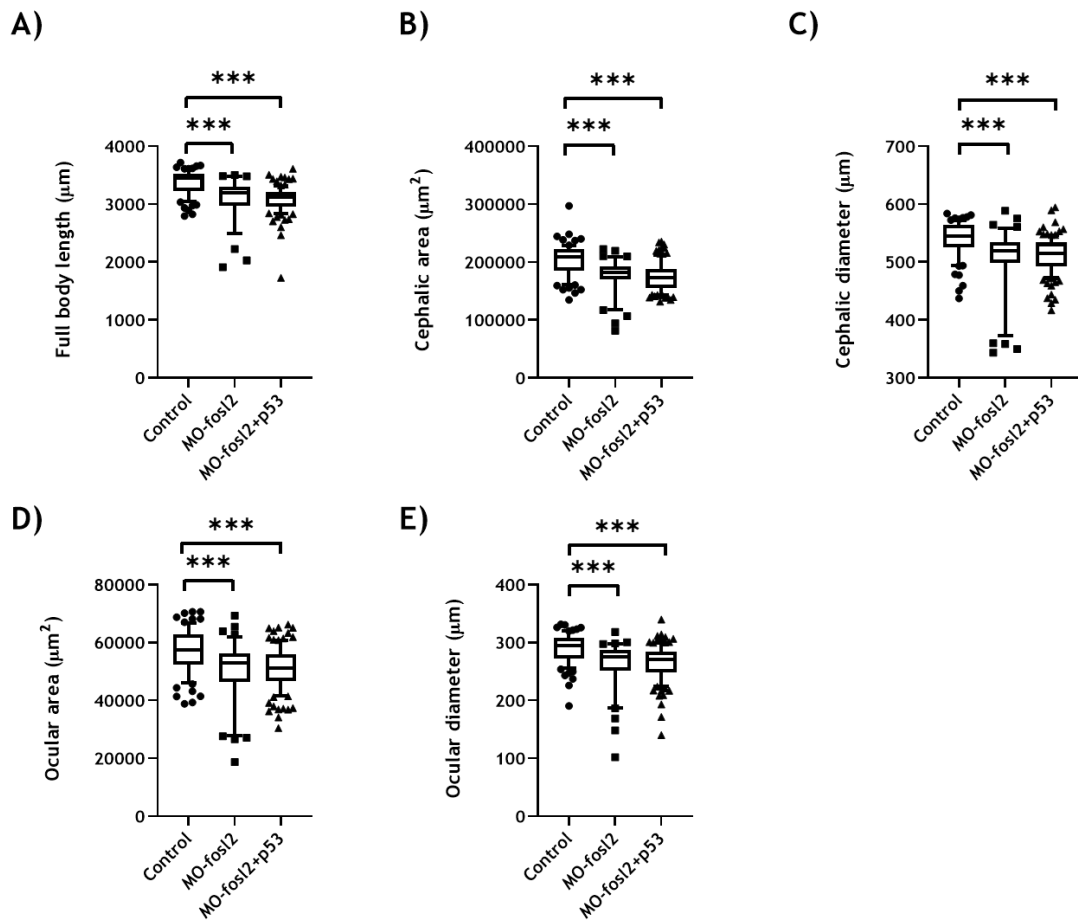


Figure 83. Bright-field images of 48 hpf zebrafish individuals. From left to right: control, treated with MO-*fosl2*, treated with both MO-*fosl2* and MO-*p53*.

Table 33 displays a summary of the different morphometric parameters analysed in the embryos, together with their respective mean and standard deviation (SD). The statistical analysis of the collected data is presented in Figure 84. The results of the statistical analysis show significant differences in the full body length between the control and morphant groups (MO-*fosl2* and MO-*fosl2*+MO-*p53*), with the former group being larger. Data analysis also indicated that control zebrafish seem to have larger heads and eyes in comparison to the morphant embryos. Also, no differences were found between the two morphant groups (MO-*fosl2* and MO-*fosl2*+MO-*p53*), which suggest that the previously mentioned observations can be associated to the specific knockdown of *fosl2*, rather than by p53-mediated neurotoxicity.

Table 33. Morphometric measurements conducted in the individuals treated with morpholines at 48 hpf.

PARAMETER	EXPERIMENTAL CONDITION	MEAN	SD
Full body length (µm)	Control	3374	218.8
	MO- <i>fosl2</i>	3071	379.8
	MO- <i>fosl2</i> + MO- <i>p53</i>	3081	233.4
Cephalic area (µm ²)	Control	203107	28906
	MO- <i>fosl2</i>	173120	32239
	MO- <i>fosl2</i> + MO- <i>p53</i>	174318	24452
Cephalic diameter (µm)	Control	539.0	31.36
	MO- <i>fosl2</i>	500.8	60.90
	MO- <i>fosl2</i> + MO- <i>p53</i>	512.4	31.29
Ocular area (µm ²)	Control	57049	7721
	MO- <i>fosl2</i>	49645	11283
	MO- <i>fosl2</i> + MO- <i>p53</i>	51153	7217
Ocular diameter (µm)	Control	289.6	26.41
	MO- <i>fosl2</i>	260.2	45.45
	MO- <i>fosl2</i> + MO- <i>p53</i>	264.5	29.37



COMPARED GROUPS	FULL BODY LENGTH	CEPHALIC AREA	CEPHALIC DIAMETER	OCULAR AREA	OCULAR DIAMETER
Control vs MO- <i>fosl2</i>	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Control vs MO- <i>fosl2</i> + MO- <i>p53</i>	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
MO- <i>fosl2</i> vs MO- <i>fosl2</i> + MO- <i>p53</i>	0.709	> 0.999	> 0.999	> 0.999	> 0.999

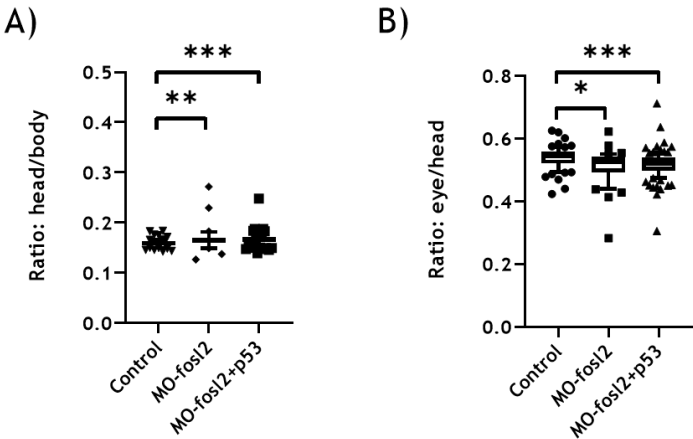
Figure 84. Morphometric analysis of the individuals treated with morpholines at 48 hpf. Analysis conducted with wild type individuals (control), individuals injected with *fosl2* morpholine (MO-*fosl2*) and individuals co-injected with *fosl2* and *p53* morpholines (MO-*fosl2* + MO-*p53*). A) Full body length. B) Cephalic area. C) Cephalic diameter. D) Ocular area. E) Ocular diameter. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

In addition to these results, two ratios of interest were also quantified: the ratio between the head and body, and the ratio between the eye and head of the embryos (Table 34). In the early stages of zebrafish development, the head is relatively larger compared to the body. However, as the animals grow, the body size increases at a faster rate than the head, causing this ratio to decrease over time. The same happens with the eye/head ratio, which decreases as the individual grows since the eye size increases at a smaller rate than the head. As can be observed in Figure 85, the head/body ratio is bigger in morphant individuals, which indicates that they have larger heads in comparison with the controls. This observation is also supported

by the eye/head ratio, which is smaller in morphants. Taken together, these results suggest that, proportionally, morphant embryos have a bigger head than their control counterparts.

Table 34. Morphometric ratios of the individuals treated with morpholines at 48 hpf.

PARAMETER	EXPERIMENTAL CONDITION	MEAN	SD
Ratio: head/body [Cephalic diameter (μm)/full body length (μm)]	Control	0.1593	0.0077
	MO- <i>fosl2</i>	0.1675	0.0239
	MO- <i>fosl2</i> + MO- <i>p53</i>	0.1672	0.0122
Ratio: eye/head [Cephalic diameter (μm)/full body length (μm)]	Control	0.5382	0.0348
	MO- <i>fosl2</i>	0.5140	0.0551
	MO- <i>fosl2</i> + MO- <i>p53</i>	0.5177	0.0433



COMPARED GROUPS	RATIO HEAD/BODY	RATIO EYE/HEAD
Control vs MO- <i>fosl2</i>	0.009	0.021
Control vs MO- <i>fosl2</i> + MO- <i>p53</i>	< 0.001	< 0.001
MO- <i>fosl2</i> vs MO- <i>fosl2</i> + MO- <i>p53</i>	0.511	> 0.999

Figure 85. Morphometric analysis of the individuals treated with morpholines at 48 hpf. Analysis conducted with wild type individuals (control), individuals injected with *fosl2* morpholine (MO-*fosl2*) and individuals co-injected with *fosl2* and *p53* morpholines (MO-*fosl2* + MO-*p53*). A) Ratio between the cephalic diameter and the full body length of the individuals. A) Ratio between the ocular diameter and the cephalic diameter of the individuals. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

In summary, morphometric analysis at 48 hpf revealed that morphant individuals have a smaller bodies, heads, and eyes, which suggest a developmental delay in comparison to control individuals of the same age. This hypothesis is also supported by the observation that, proportionally, morphants have larger heads, which are characteristic of earlier developmental stages in this animal model.

4.2.3.2. Cartilage staining of *fosl2* morphants.

The presence of alterations in the formation of the jaw were studied in 120 hpf zebrafish by measuring the ceratohyal angle and the palatoquadrate length (Figure 86). The analysis of these parameters revealed that *fosl2* knockdown does not have major repercussions in the cartilage formation in zebrafish. Nonetheless, the results show a reduction in the palatoquadrate length of MO-*fosl2* + MO-*p53* individuals in comparison to MO-*fosl2* zebrafish,

but this small alteration could be linked to the knockdown of p53 since no other significant differences were found among the experimental groups.

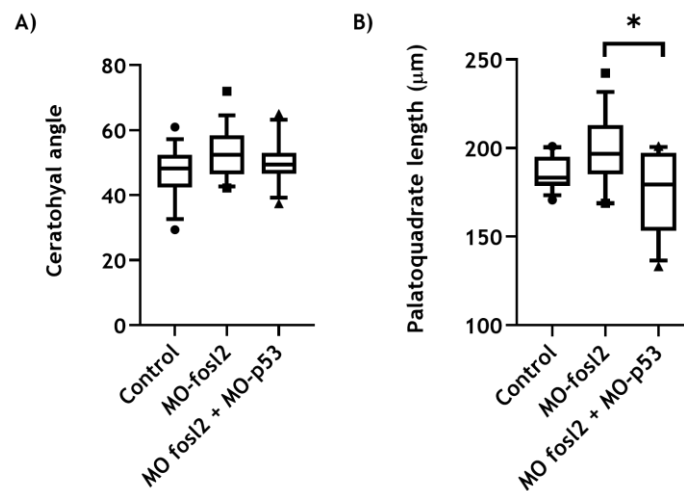


Figure 86. Cartilage staining of *fosl2* morphants. A) Ceratohyal angle. B) Palatoquadrate length. Data represented as 10-90 percentile. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

4.2.3.3. Behavioural response of the *fosl2* morphants using the light-dark test.

The light-dark test is frequently used in the zebrafish research field to study anxiety-like responses mediated by cyclic alterations in the illumination. Light to dark changes trigger the anxiety-like mechanisms in zebrafish, thus increasing the locomotor activity of the individuals. On the other hand, dark to light changes have the opposite effect.

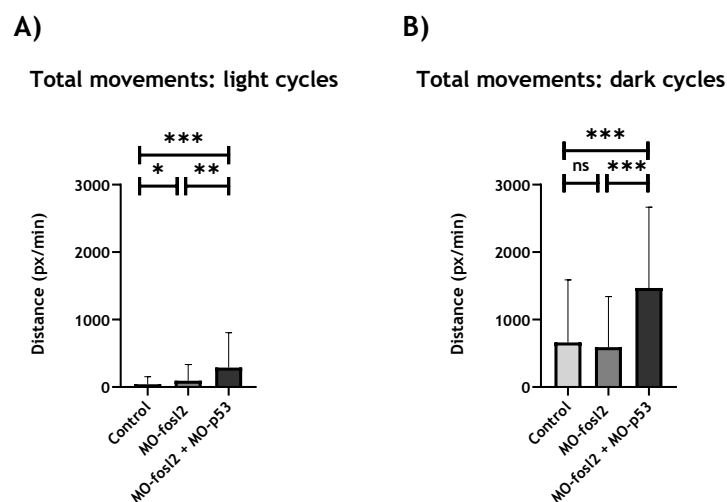


Figure 87. General representation of the movements of the individuals at 120 hpf in light and dark cycles. Analysis conducted with wild type individuals (control), individuals injected with *fosl2* morpholine (MO-fosl2) and individuals co-injected with *fosl2* and *p53* morpholines (MO-fosl2 + MO-p53). A) Movements detected in light cycles. B) Movements detected in dark cycles. Data represented as mean + SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

Here, the light-dark test was used to study the anxiety response in *fosl2* morphant zebrafish (MO-*fosl2* and MO-*fosl2* + MO-*p53*) to study whether the knockdown of the gene had consequences at a behavioural level.

A tracking system was used to quantify the distance travelled by each zebrafish individual at 120 hfp in light and dark cycles, which is represented as pixels travelled each minute. The result of this test revealed that the anxiety-like response is significantly higher in zebrafish treated with MO-*fosl2* + MO-*p53* than in the other experimental conditions. In addition, no differences were found when comparing the control and MO-*fosl2* groups when the cycles were individually analysed. In addition, the total movements in light and dark cycles were analysed, which revealed that MO-*fosl2* morphants are more active than the controls in light cycles, but less than the MO-*fosl2* + MO-*p53* treated individuals (Figure 87, Figure 88).

Taken together, the light-dark test results suggest that the locomotion alterations might be a consequence of the knockdown of *p53*, rather than an effect of the knockdown of *fosl2*.

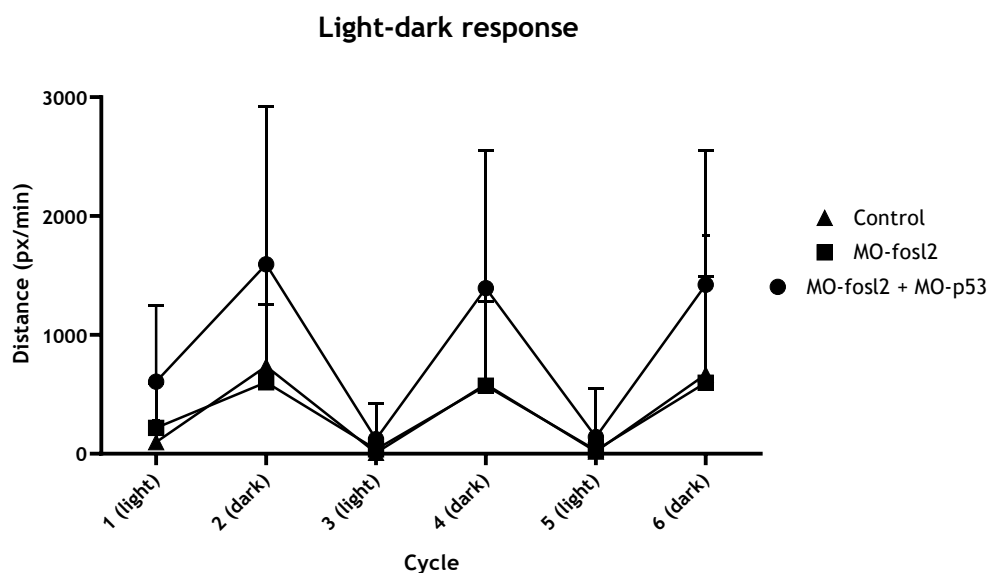


Figure 88. Light-dark response test of *fosl2* morphant zebrafish. Analysis conducted with wild type individuals (control), individuals injected with *fosl2* morpholine (MO-*fosl2*) and individuals co-injected with *fosl2* and *p53* morpholines (MO-*fosl2* + MO-*p53*). Movements were measured as distance (in pixels) travelled per minute. Data represented as mean + SD.

Table 35. P-values ($\alpha = 0.05$) obtained in the statistical analyses conducted with data obtained from light-dark test. Statistically significant values are indicated with the symbol (*).

EXPERIMENTAL GROUP	CYCLE 1 (LIGHT)	CYCLE 2 (DARK)	CYCLE 3 (LIGHT)	CYCLE 4 (DARK)	CYCLE 5 (LIGHT)	CYCLE 6 (DARK)
Control vs MO- <i>fosl2</i>	0.148	> 0.999	0.410	0.999	0.815	> 0.999
Control vs MO- <i>fosl2</i> + MO- <i>p53</i>	< 0.001*	0.019*	0.001*	0.005*	0.008*	0.012*
MO- <i>fosl2</i> vs MO- <i>fosl2</i> + MO- <i>p53</i>	0.024*	0.001*	0.091	0.007*	0.123	0.006*

4.3. GENERATION AND CHARACTERIZATION OF *IN VITRO* AND *IN VIVO* MODELS TO STUDY THE FUNCTION OF *TAOK1*.

The objective of this section is to describe the generation and characterization of a mutant hiPSCs model, as well as an *in vivo* zebrafish model, to study the function of *TAOK1* and its association with neurodevelopmental alterations and ASD.

4.3.1. Generation of a knockout model for *TAOK1* *in vitro*.

This section is going to be dedicated to the generation and characterization of a knockout hiPSCs cell line to study the function of *TAOK1* and its role in neurodevelopment and its alterations, such as ASD.

4.3.1.1. CRISPR-Cas9 guide design, cloning into suitable vectors, and transfection.

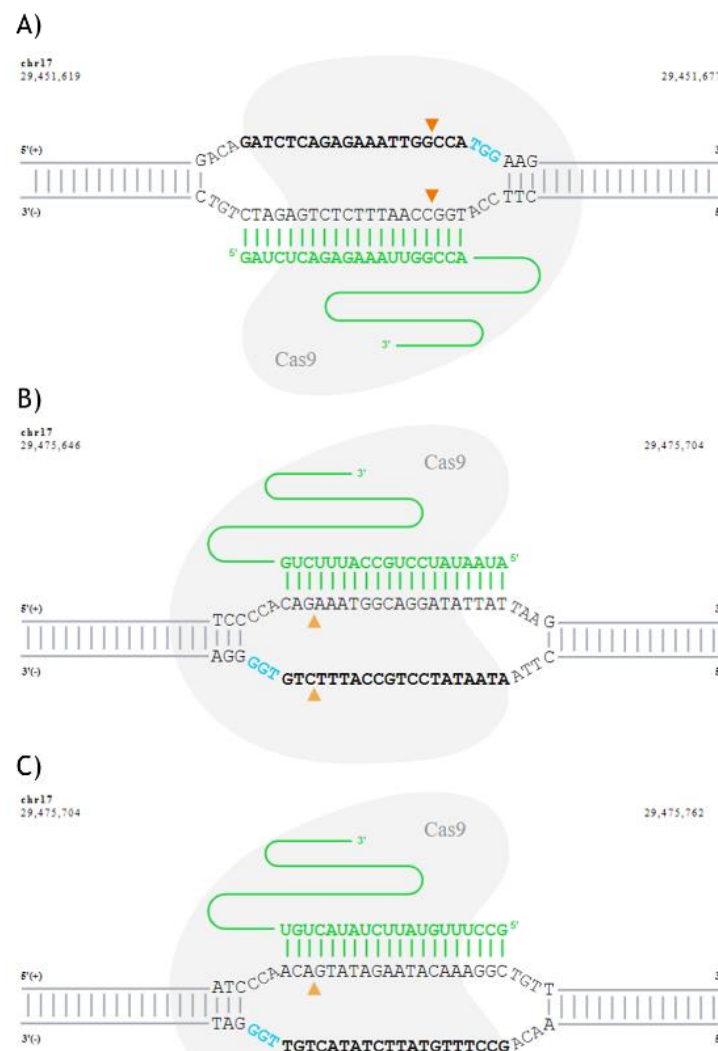


Figure 89. Schematic representation of the sgRNAs designed for the knockout of *TAOK1* on hiPSCs using the CRISPR-Cas9 system. A) *TAOK1*_sgRNA_1. B) *TAOK1*_sgRNA_2. C) *TAOK1*_sgRNA_3. Image generated by the author using Synthego's Verify Guide Design webtool v1.3²⁸².

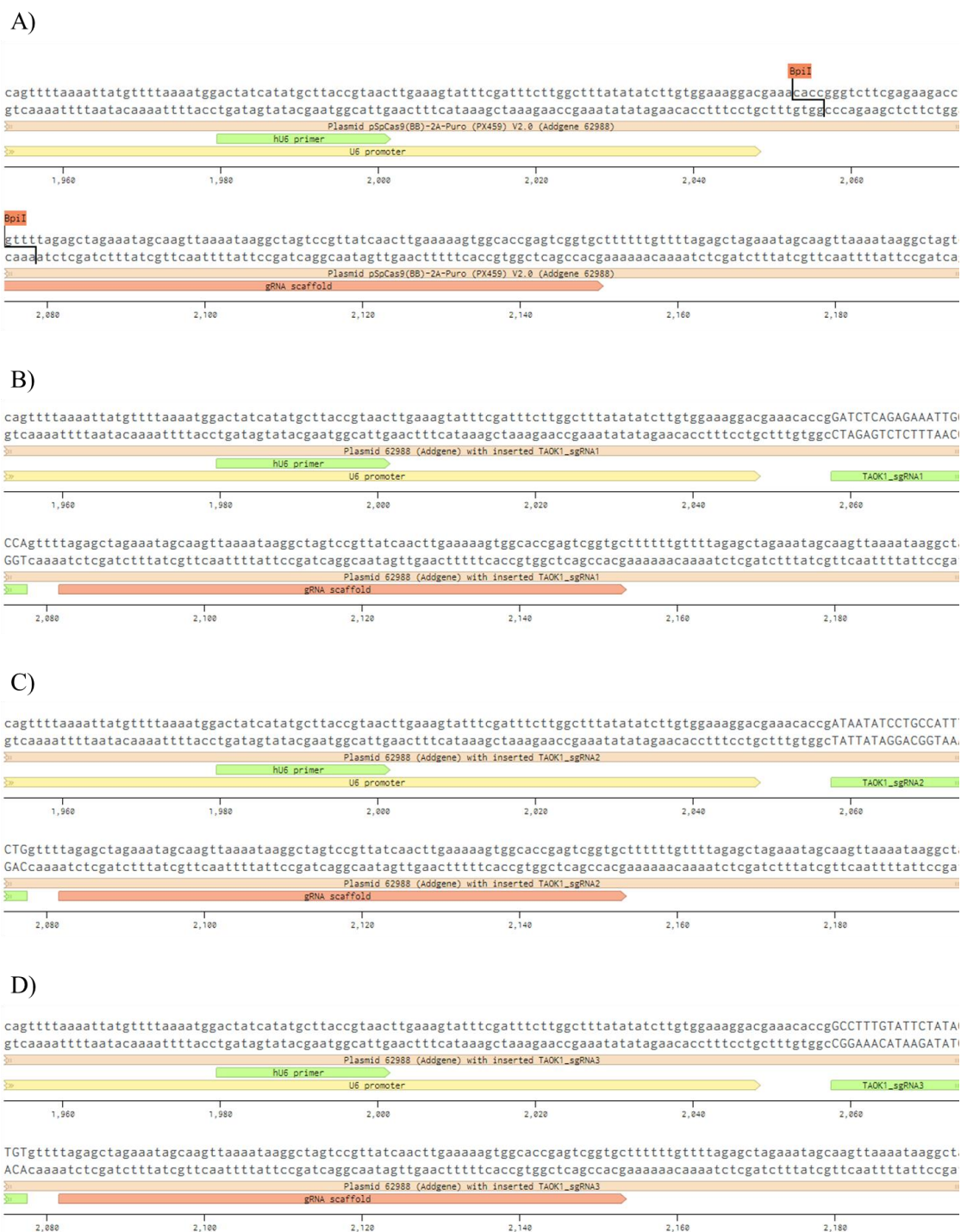


Figure 90. Representation of the plasmid pSpCas9 (BB)-2A-Puro (Addgene #62988) before and after the insertion of the guides to knockout *BTBD8*. **A)** Plasmid 62988 without modifications. **B)** Plasmid 62988 after the insertion of the *BTBD8_sgRNA1*. **C)** Plasmid 62988 after the insertion of the *BTBD8_sgRNA2*. **D)** Plasmid 62988 after the insertion of the *BTBD8_sgRNA3*. Images generated by the author using Benchling webtool.

The first step to knockout *TAOK1* in hiPSCs was to design suitable guides to target the gene of interest. Initially, CRISPR-ERA webtool²⁸⁰ was applied for that purpose, giving a total amount of 304 potential guides to knockout *TAOK1* in *Homo sapiens*. Next, the most appropriate guides to be expressed with the U6 promoter were selected. The resulting top 50

guides were filtered attending to the target localization, prioritizing guides that bind to exons 1 - 10 in all the protein transcripts. Additionally, only guides with a positive E+S CRISPR-ERA score were considered, in order to rule out the less efficient guides with potential off-targets in the genome. Two of the guides passed all these filters and were further analysed using Synthego's Verify Guide Design webtool²⁸². This second algorithm also considered them as guides that target common exons among all transcripts, with high efficiency, and minimal off-targets (Figure 89, A and C). Since we wanted to try at least three different guides to maximize the probability of successfully knockout *TAOK1*, we also run Synthego's Knockout Guide Design v1.3²⁸¹ and selected the recommended guide to knockout the gene as well (Figure 89, B).

Afterwards, guides were cloned separately into the plasmid vector pSpCas9 (BB)-2A-Puro (Addgene #62988) with the restriction enzyme *Bpi*I. A model of the expected results was conducted *in silico* as indicated in Figure 90, on which can be observed the seamless insertion of each guide between the U6 promoter and the sgRNA scaffold (tracrRNA). The predicted results were used as a reference to align the sequences of the cloned plasmids for each guide, to discard those that include alterations (Figure 91). In Figure 91 can be observed that some of the analysed plasmids include modifications of their sequence that could alter the functionality of the plasmids, and thus, affect the CRISPR-Cas9 efficiency. However, it was possible to identify plasmids without alterations as well, that could be transfected into hiPSCs to knockout *TAOK1*. The plasmids chosen for that purpose were the ones represented in Figure 91 with the names *TAOK1_sgRNA1_sample1*, *TAOK1_sgRNA2_sample 3* and *TAOK1_sgRNA3_sample 5*.

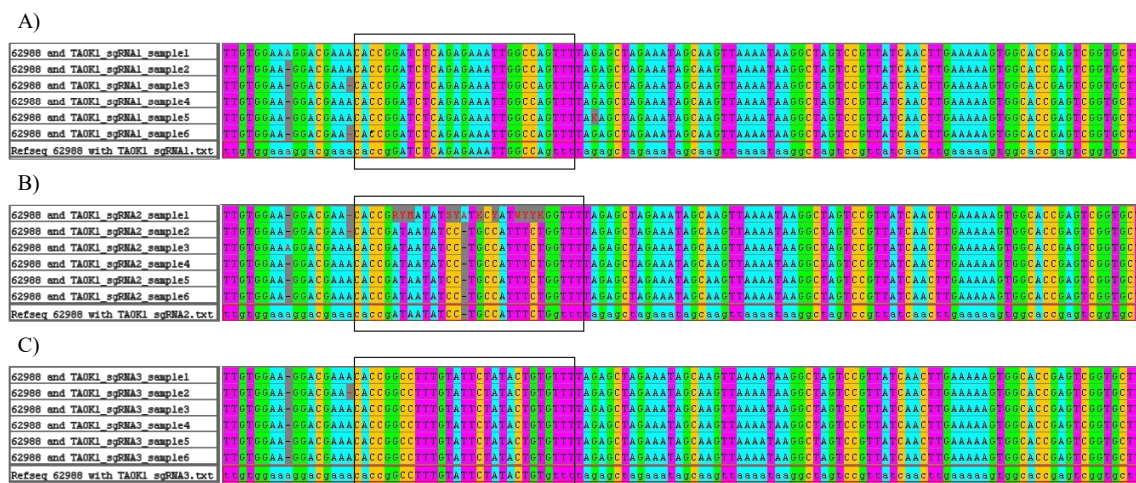


Figure 91. Verification of the correct insertion of the knockout guides for *TAOK1* in the plasmid pSpCas9 (BB)-2A-Puro (Addgene #62988) by Sanger sequencing. A) Plasmids cloned with *TAOK1_sgRNA1* (samples 1-6) aligned with the predicted sequence expected after the insertion (indicated as “Refseq” in the image). B) Plasmids cloned with *TAOK1_sgRNA2* (samples 1-6) aligned with the predicted sequence expected after the insertion (indicated as “Refseq” in the image). C) Plasmids cloned with *TAOK1_sgRNA3* (samples 1-6) aligned with the predicted sequence expected after the insertion (indicated as “Refseq” in the image). Black rectangles indicate the position of the inserted guide. Image created by the author in Codon Code Aligner v10.0.2.

The previously mentioned plasmids were transfected, separately, into hiPSCs. After selecting the cells for two days with puromycin (1 μ g/mL), a sample of the transfected populations was taken to extract DNA and analyse the knockout efficiency of each guide.

4.3.1.2. Guide efficiency analysis.

As it was previously mentioned, three guides were designed to knockout *TAOK1*, all of them with predicted high efficiencies *in silico*. However, these predictions not always translate to the experimental findings. With the objective to avoid growing clones transfected with low-performance guides, and focusing the efforts on the most successful ones, a guide efficiency analysis was conducted.

First of all, the ICE CRISPR Analysis tool²⁸⁴ was applied to calculate the efficiency of each guide. This tool compares sequencing data from the pool of transfected cells with their WT isogenic control, inferring the editing efficiency by analysing differences in the target region of each guide. In Figure 92, the traces tabs of the analysed samples are represented. In that figure, a clean Sanger sequence for the WT samples (control) can be observed in all the target sites for the different guides, which is an indicative of the quality of the data, a key point to a successful ICE analysis.

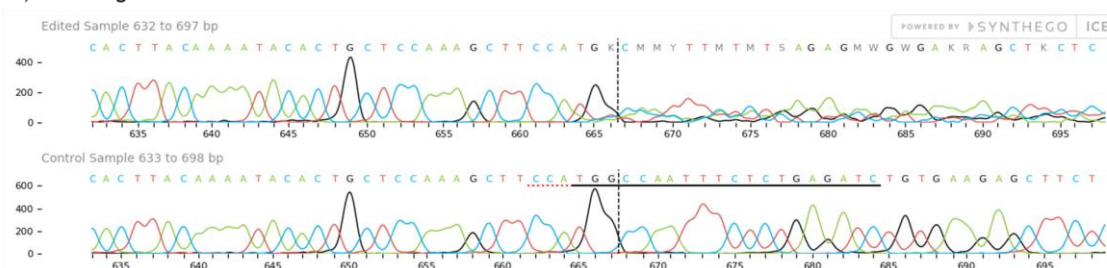
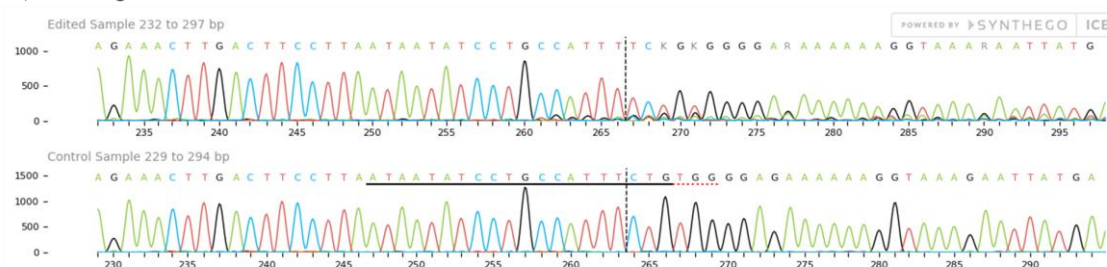
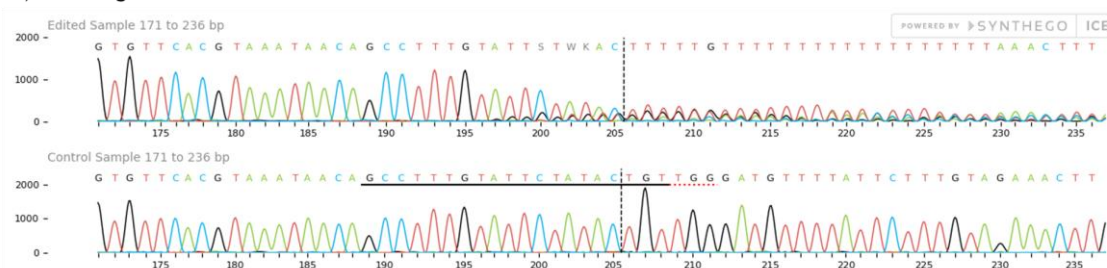
A) *TAOK1* sgRNA1B) *TAOK1* sgRNA2C) *TAOK1* sgRNA3

Figure 92. Traces tab of the ICE analysis conducted for the guides designed to knockout *TAOK1*. Sanger sequencing data of the edited pool of cells is compared to the wild type isogenic control for the three different guides. A) *TAOK1*_sgRNA1. B) *TAOK1*_sgRNA2. C) *TAOK1*_sgRNA3. The target sequence for each guide is underlined in black, while the PAM sequence is indicated underlined with red dots.

The discordance plot in Figure 93 indicates the level of disagreement between the WT and the edited samples, showing a good overlay until the CRISPR-Cas9 cut site, also known as interference region. In the interference region, the level of disagreement between both sequences increases, which is an indicative of the activity of the CRISPR-Cas9 system. In that

same figure, it is also represented a histogram of the predicted indel sizes in the sample and their abundance, along with the predicted editing efficiency and the Pearson correlation coefficient (R^2). For all the guides tested, the R^2 of the model is higher than 0.8, so all of them can be considered reliable. Moreover, all the designed guides to knockout *TAOK1* have a high editing efficiency ($> 70\%$), being the higher scoring guide *TAOK1*_sgRNA2 (89 %), followed by *TAOK1*_sgRNA3 (80 %), and *TAOK1*_sgRNA1 (76 %). *TAOK1*_sgRNA1 was the most interesting guide for the purposes of this thesis since it targets *TAOK1* exon 2, rather than exon 4 like *TAOK1*_sgRNA2 and *TAOK1*_sgRNA3. Despite not having the highest predicted efficiency, the fact that *TAOK1*_sgRNA1 targeted an earlier coding exon could increase the probability of obtaining a full knockout in hiPSCs.

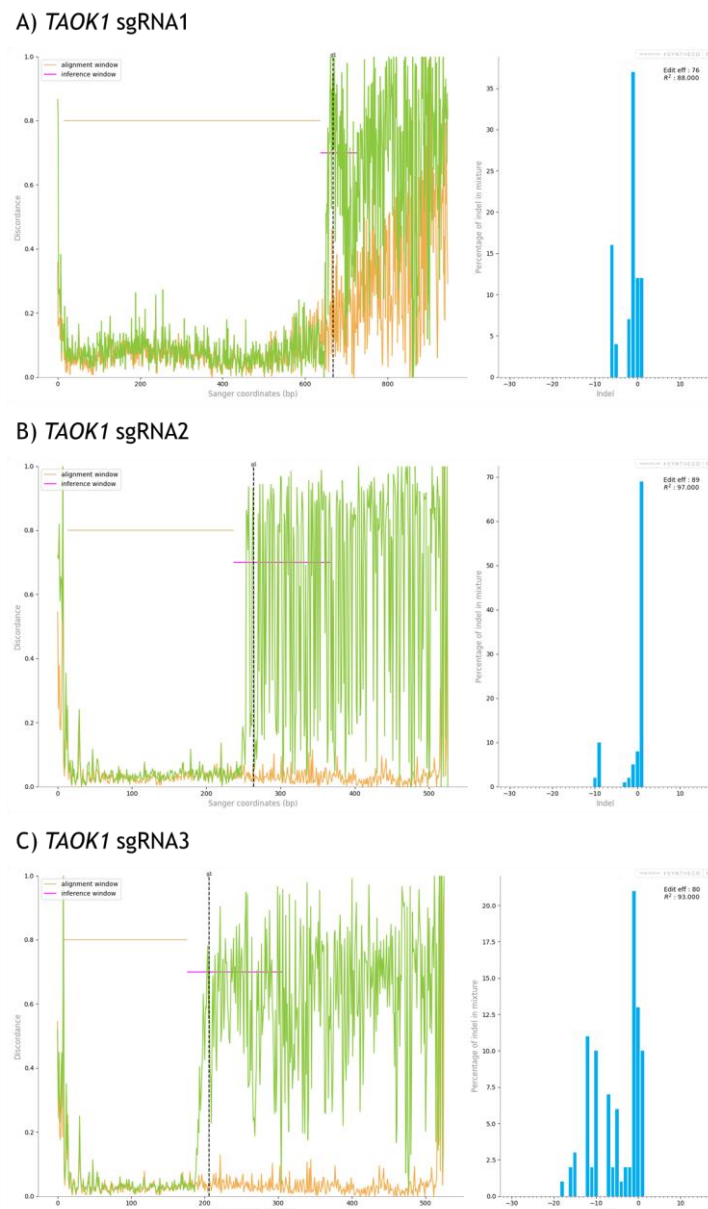


Figure 93. ICE analysis results for the knockout guides for *TAOK1*. Disagreement plot (left) with the wild type sequence in orange and the edited samples in green. Histogram of the frequencies of indels in the edited samples (right). R^2 (Pearson correlation coefficient) indicates the model fit value. A) *TAOK1*_sgRNA1, R^2 0.88, editing efficiency 76 %. B) *TAOK1*_sgRNA2, R^2 0.97, editing efficiency 89 %. C) *TAOK1*_sgRNA3 R^2 0.93, editing efficiency 80%.

A second approach was followed to double-check the efficiency of *TAOK1*_sgRNA1, before definitively selecting the pool of cells edited with that guide to establish a homogeneous cell line to work with. The second method to test guide efficiency that was applied was based on the direct identification of alleles in the pool of edited cells, which was possible by isolating them with bacterial cloning step. The results obtained with this approach, represented in Figure 94 and Table 36, support the high efficiency of the *TAOK1*_sgRNA1, since all the identified alleles in the pool of edited cells had mutations in the target region for the guide. These results further support the selection of the pool of *TAOK1*_sgRNA1 edited cells for the isolation of an edited knockout hiPSCs cell line for *TAOK1*.



Figure 94. Predicted efficiency of the guide *TAOK1*_sgRNA1 (in yellow). Alignment of the reference sequence of *TAOK1* (GRCh38), on top, with the alleles isolated by TOPO cloning from the bulk of edited cells. Image created by the author using the SnapGene software.

Table 36. Sequence alterations identified by TOPO cloning of the cells treated with *TAOK1*_sgRNA1.

BARCODE	TARGETED GENE	EXON	GUIDE	BACTERIAL CLONE	SEQUENCE ALTERATIONS
EF70647351	<i>TAOK1</i>	2	<i>TAOK1</i> _sgRNA1	1	Deletion of 12 nts
EF70647352	<i>TAOK1</i>	2	<i>TAOK1</i> _sgRNA1	2	Deletion of 9 nts
EF70647353	<i>TAOK1</i>	2	<i>TAOK1</i> _sgRNA1	3	Insertion of 6 nts
EF70647354	<i>TAOK1</i>	2	<i>TAOK1</i> _sgRNA1	4	Deletion of 59 nts
EF70647355	<i>TAOK1</i>	2	<i>TAOK1</i> _sgRNA1	5	Deletion of 1 nt and several base changes
EF70647356	<i>TAOK1</i>	2	<i>TAOK1</i> _sgRNA1	6	Deletion of 1 nt

4.3.2. Characterization of the generated clones.

Once the efficiency of the guide was verified, the pool of cells edited *TAOK1*_sgRNA1 were seeded at low density, so that clones derived from single cells could be isolated and characterized, as described in the following subheadings.

4.3.2.1. Allele identification.

The first step in the characterization of the isolated single cell clones was to identify the presence of mutations in *TAOK1*. In this case, 29 clones were isolated from the pool of edited cells with *TAOK1*_sgRNA1. One of the 29 clones (clone 19) was discarded before the genetic characterization since the cells exhibited excessive spontaneous differentiation. The remaining 28 clones were sequenced with the objective of identifying the ones without genetic modifications from the edited ones. The sequencing results indicate that 2 clones (clone 24 and clone 26) showed no signs of genetic modification, since the WT sequence was clean, as indicated in Figure 95. The remaining 26 clones showed signs of being potentially edited. It

was noticeable that clones 10 (Figure 96), 18 (Figure 97) and 28 (Figure 98) could potentially have homozygous mutations in *TAOK1*, since the Sanger sequencing results clearly display indels (1 nt insertion, 3 nts deletion, and 4 nts deletion, respectively). However, this could also be observed if just one allele had the detected mutation, while the other is affected by a large deletion. To avoid this possibility, these clones were not used for further analysis on this thesis.

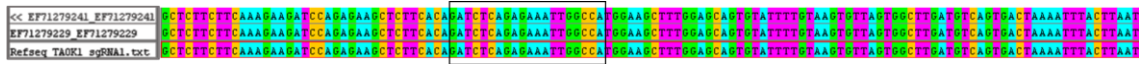


Figure 95. Clone 24 isolated from the edition with *TAOK1_sgRNA1*. Alignment with the reference sequence of *TAOK1* (GRCh38) with the Sanger sequences obtained from the clone of interest.



Figure 96. Clone 10 isolated from the edition with *TAOK1_sgRNA1*. Alignment with the reference sequence of *TAOK1* (GRCh38) with the Sanger sequences obtained from the clone of interest.

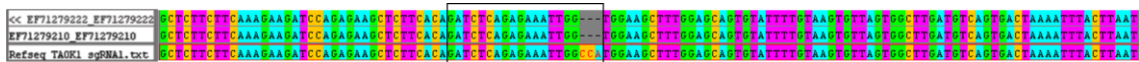


Figure 97. Clone 18 isolated from the edition with *TAOK1_sgRNA1*. Alignment with the reference sequence of *TAOK1* (GRCh38) with the Sanger sequences obtained from the clone of interest.

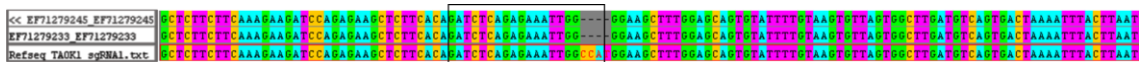
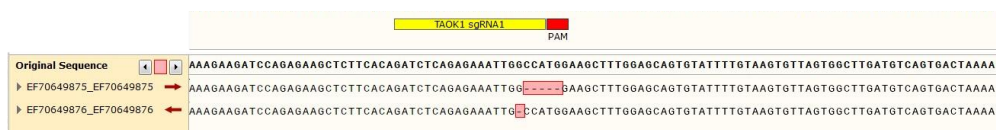
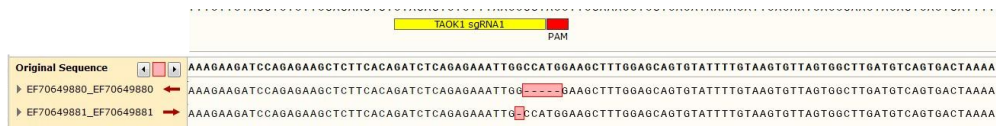


Figure 98. Clone 28 isolated from the edition with *TAOK1_sgRNA1*. Alignment with the reference sequence of *TAOK1* (GRCh38) with the Sanger sequences obtained from the clone of interest.

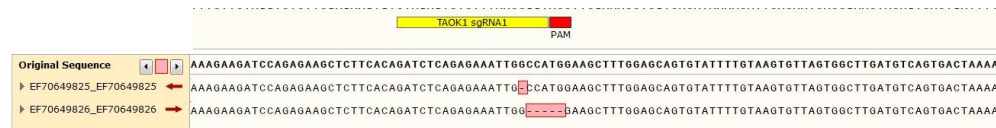
A) *TAOK1* sgRNA1 clone 4



B) *TAOK1* sgRNA1 clone 6



C) *TAOK1* sgRNA1 clone 7



D) *TAOK1* sgRNA1 clone 8

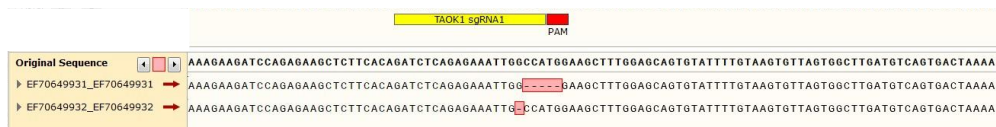


Figure 99. Alignment with the reference sequence of *TAOK1* (GRCh38) with the Sanger sequences obtained from several clones of interest after TOPO cloning. A) Alleles present in the clone 4 isolated from the edition with *TAOK1_sgRNA1*. B) Alleles present in the clone 6 isolated from the edition with *TAOK1_sgRNA1*. C) Alleles present in the clone 7 isolated from the edition with *TAOK1_sgRNA1*. D) Alleles present in the clone 8 isolated from the edition with *TAOK1_sgRNA1*.

Of the 23 potentially edited clones left, bacterial cloning was applied to 4 of them to precisely identify the allele composition of the cell population (Figure 99). Clones 4, 6, 7 and 8 had heterozygous mutations in *TAOK1*, all of them with two frameshift mutations that cause

early stop codons to appear in *TAOK1* coding sequence: a deletion of 5 nts (NC_000017.11:g.29451653_29451657delCCATG) in one allele, and a 1 nt deletion in the other (NC_000017.11:g.29451652delG). The 5 nts frameshift mutation is expected to generate a 49 aa protein (NP_065842.1:p.His36Lysfs*14), while the 1 nt frameshift mutation is predicted to encode a 82 aa protein (NP_065842.1:p.Gly35Alafs*48), both of them represented in Figure 100. Those altered versions of *TAOK1* have early stop codons inside the protein kinase domain of the protein, thus potentially knocking out its activity effectively.

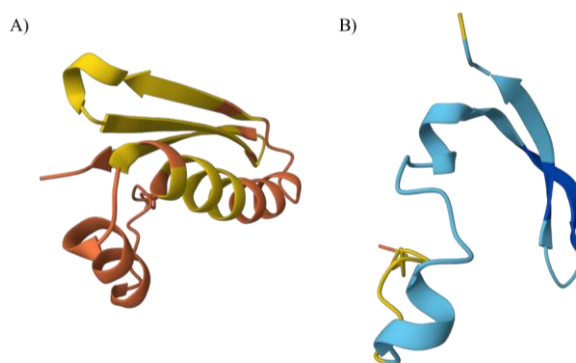


Figure 100. Predicted structures of the proteins generated by the mutant *TAOK1* alleles. A) Protein NP_065842.1:p.Gly35Alafs*48 derived from NC_000017.11:g.29451652delG. B) Protein NP_065842.1:p.His36Lysfs*14 derived from NC_000017.11:g.29451653_29451657delCCATG. Image generated using AlphaFold Monomer v2.0 pipeline^{214,215}.

In the light of these results, *TAOK1*_sgRNA1_clone 8 was selected to be further characterized, keeping the other potentially knockout clones as a backup in case this did not maintain stem-cell properties. That clone was re-named as Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1*-i2, following the guidelines of the *Banco Nacional de Líneas Celulares* and the *Instituto de Salud Carlos III*²⁸⁵.

4.3.2.2. Karyotype analysis of the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1*-i2.

The karyotype of the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1*-i2 was examined. A total of 21 metaphases were analysed, and each exhibited a normal pattern in terms of chromosome number and structure with the karyotype resolution. Specifically, structural abnormalities and chromosomal fragility, such as gaps and breaks, have been ruled out on this analysis. In Figure 101 it is displayed one of the analysed metaphases of this cell line.

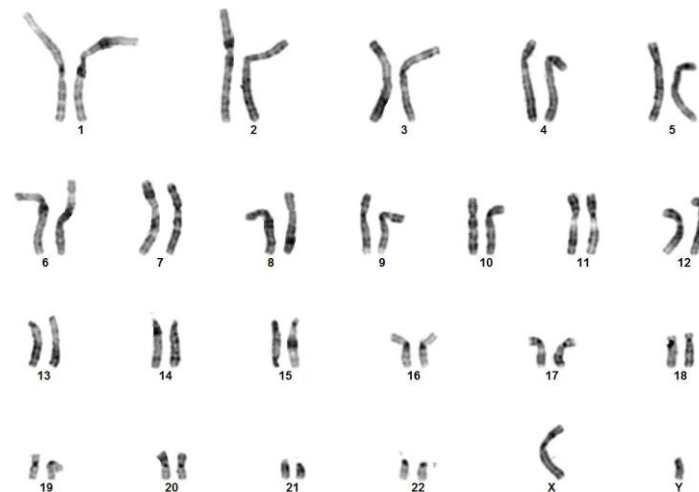


Figure 101. Karyotype of the cell line Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2.

4.3.2.3. Pluripotency marker expression in the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2.

The next step in the characterization of the edited hiPSC line Ctrl.BS PBiPs37-Sv4F-1-TAOK1-i2 was to confirm the expression of pluripotency markers. Here, the expression of *NANOG*, *POU5F1*, and *SOX2* was verified by RT-qPCR (Figure 102), revealing no significant differences in comparison with the isogenic control Ctrl.BS-PBiPs37-Sv4F-1.

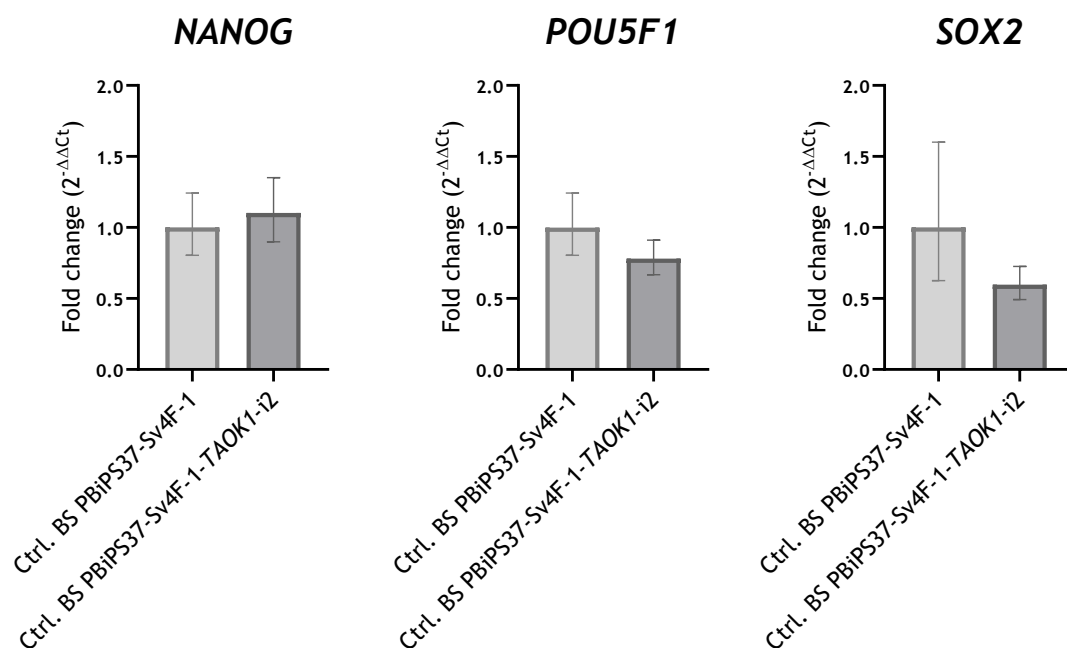


Figure 102. RT-qPCR for pluripotency markers in the knockout cell line Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2 compared to the isogenic wild type cell line Ctrl.BS-PBiPs37-Sv4F-1. Data represented as the geometric mean of the fold change and the geometric SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

Furthermore, the characterization was extended also by interrogating the expression of pluripotency markers at the protein level, where the stem cell markers NANOG, SOX2, and TRA-1-81 (Figure 103, Figure 104, Figure 105, Figure 106) were successfully detected by immunofluorescence. These findings provide additional robust evidence of the pluripotent nature of the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1*-i2.

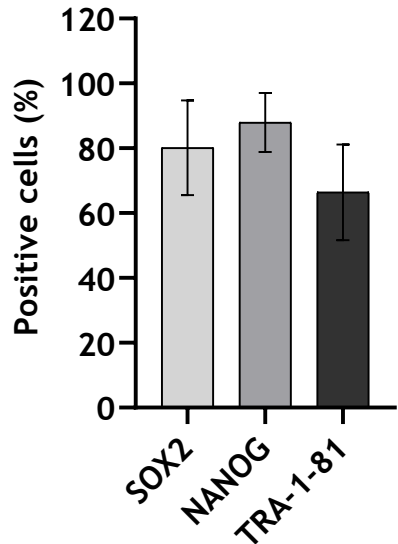


Figure 103. Quantification of immunofluorescence images for the pluripotency markers SOX2, NANOG, and TRA-1-81.

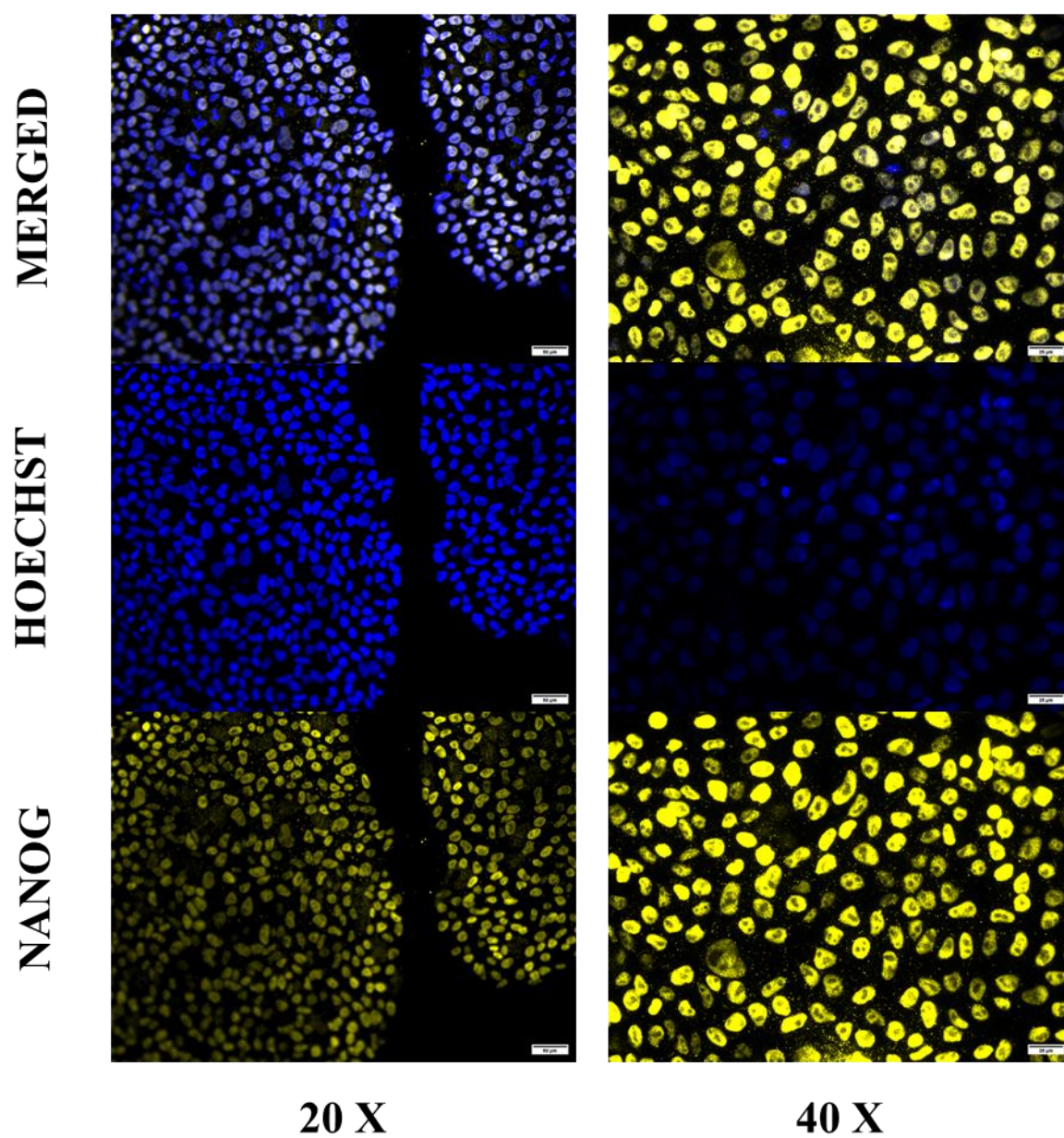


Figure 104. Immunofluorescence staining of the pluripotency marker NANOG in the cell line Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2. Nuclear staining was conducted with Hoechst.

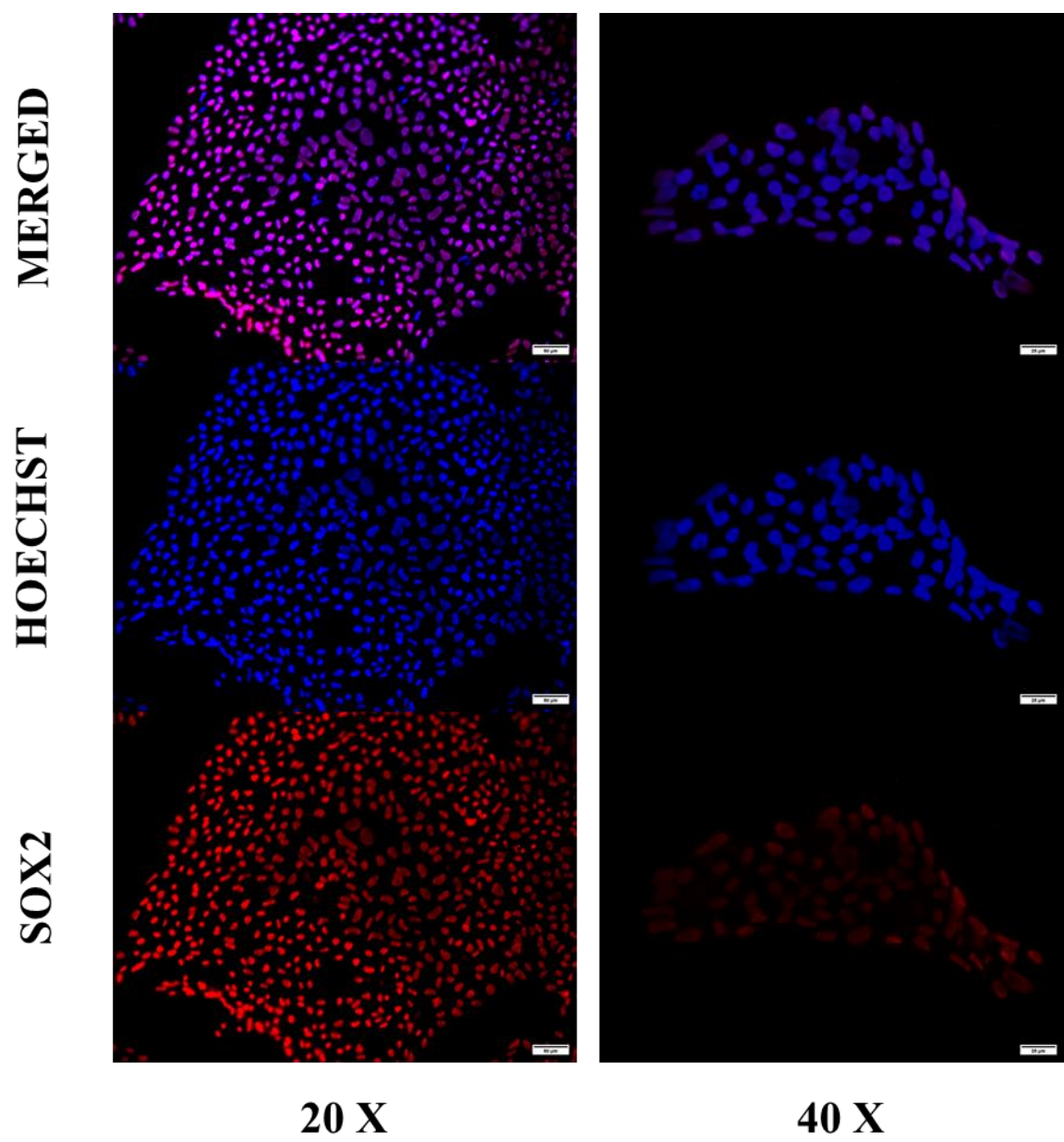


Figure 105. Immunofluorescence staining of the pluripotency marker SOX2 in the cell line Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2. Nuclear staining was conducted with Hoechst.

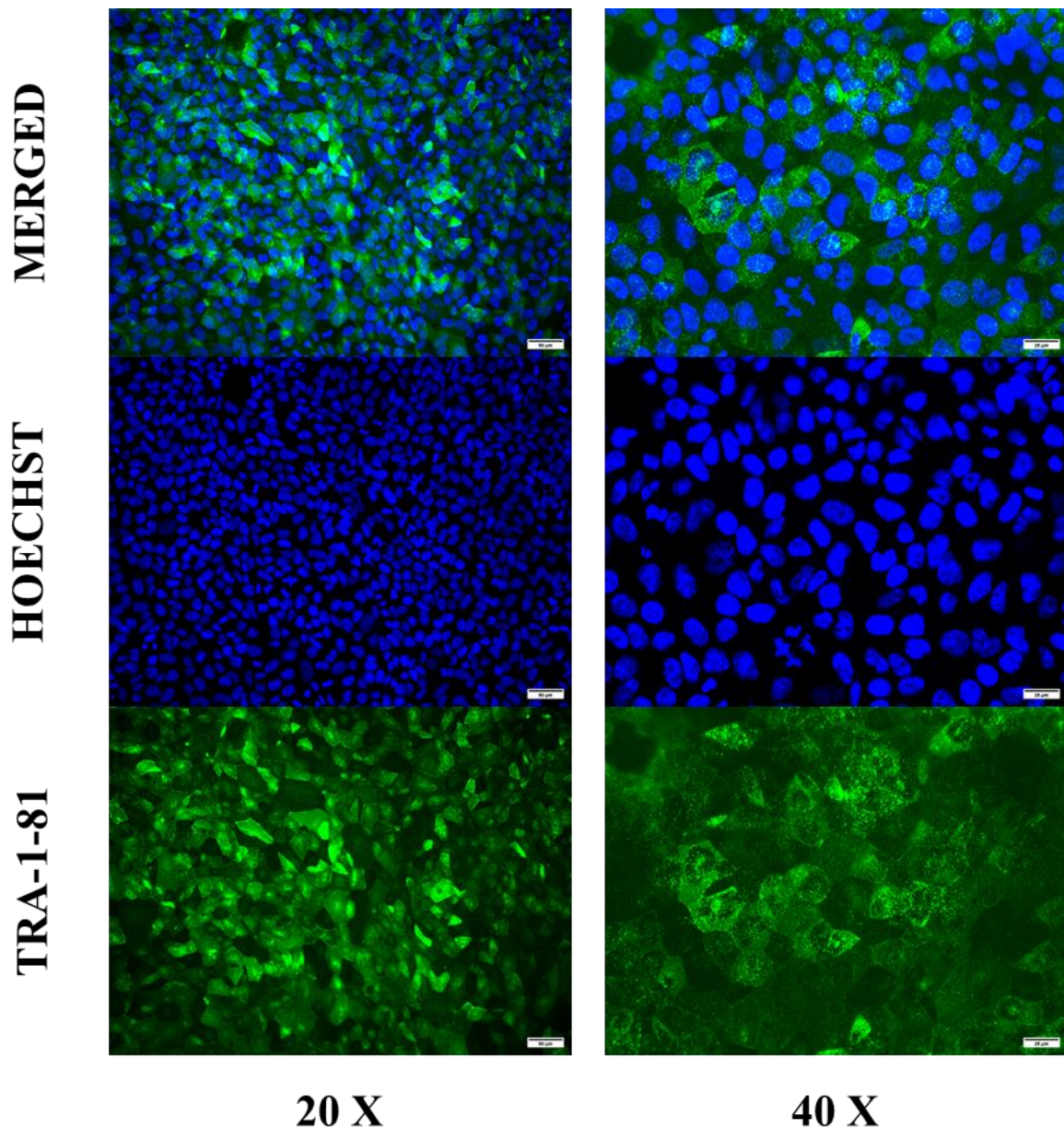


Figure 106. Immunofluorescence staining of the pluripotency marker TRA-1-81 in the cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2*. Nuclear staining was conducted with Hoechst.

4.3.2.4. Trilineage differentiation of Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2* edited cells.

The characterization of the edited hiPSC line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2* extended to evaluating its pluripotency by inducing differentiation of hiPSCs into the three germ layers (ectoderm, mesoderm, and endoderm).

Brightfield images captured at the initiation and conclusion of the differentiation protocol for each germ layer are presented in Figure 107. As observed there, no differences are noticeable between the *TAOK1*-mutant cells and their isogenic WT control for either germ layer. In addition to this, the expression of germ-layer specific markers was analysed by RT-qPCR. As represented in Figure 108, *NCAM1* was successfully detected in mesoderm-differentiated cells from both cell lines, and similar outcomes were observed for the ectoderm (*NES*) and endoderm markers (*AFP*).

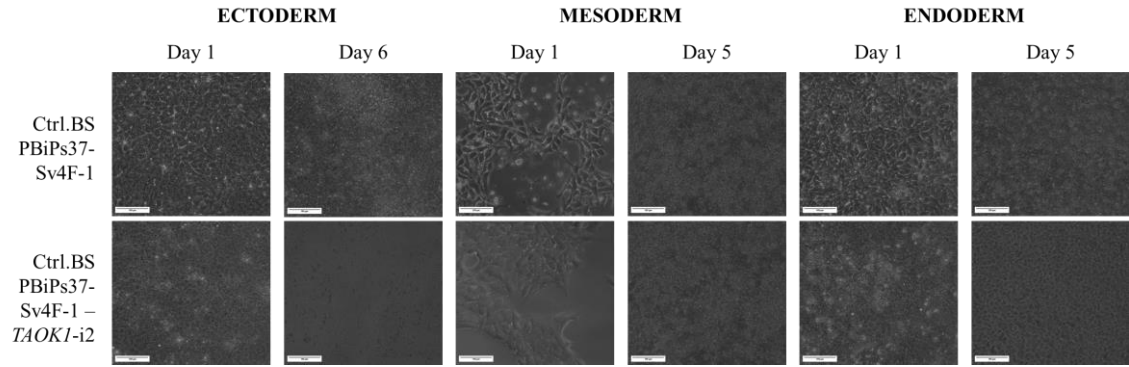


Figure 107. Brightfield images of the three germ-layer differentiation conducted using Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1*-i2 edited cells and their isogenic control (Ctrl.BS-PBiPs37-Sv4F-1).

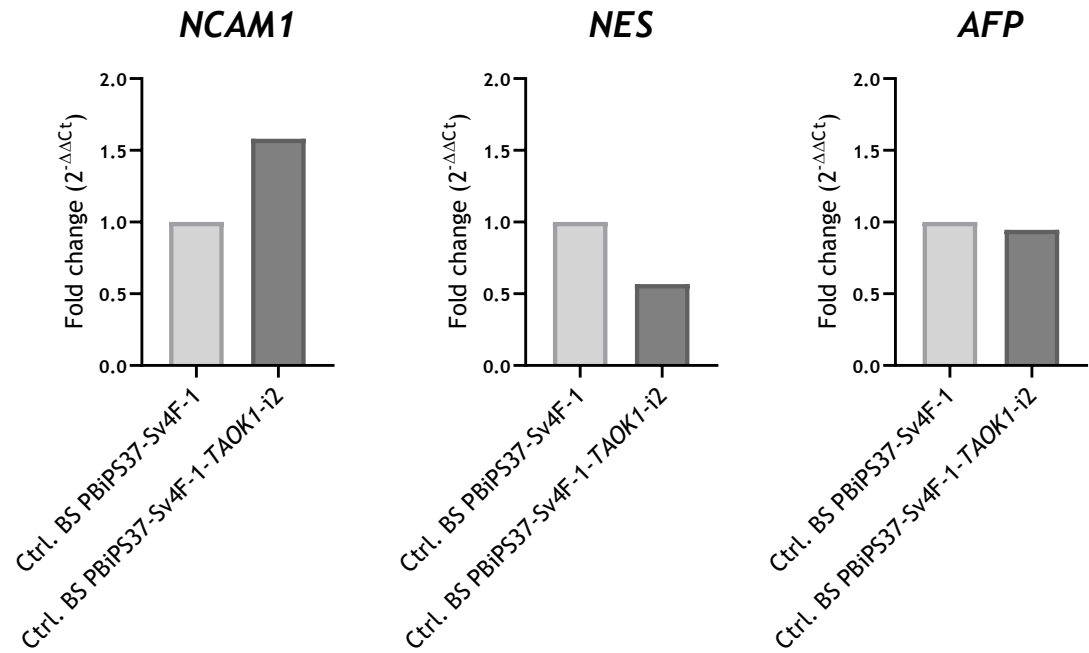


Figure 108. RT-qPCR for germ layer markers in the cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1*-i2 compared to the isogenic wild type cell line Ctrl.BS-PBiPs37-Sv4F-1. *NCAM1* expression was tested in mesoderm differentiated cells. *NES* expression was tested in ectoderm differentiated cells. *AFP* expression was tested in endoderm differentiated cells.

4.3.2.5. Off-target assessment in the Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1*-i2 edited line.

The characterization of the cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1*-i2 also included the identification of off-target genome editing events. According to the CasOFFinder⁶⁸ and Synthego's Verify Guide Desing webtools²⁸², the guide *TAOK1*_sgRNA1 could be able to target 18 positions in the human genome with up to 3 mismatches, all of them collected in Table 37. To identify alterations in any of those positions, primer pairs were designed to sequence them (Table 38).

Table 37. Predicted off-target regions for *TAOK1*_sgRNA1 identified using the CasOFFinder and Synthego's Verify Guide Desing webtools. Mismatches between the off-target region and the sgRNA sequence are indicated in lowercase.

CODE	MISMATCH	OFF-TARGET SEQUENCE	CUT SITE	GENE
OFF_T_1	2	GATCTCAGAGAAATgaGCCAAGG	chr9:86805509	-
OFF_T_2	3	GATCTCAGAGAAAcTgtCaACGG	chr8:35439333	<i>UNC5D</i>
OFF_T_3	3	GATCTCAtgGAAAcTGGCCAGGG	chr5:134146580	<i>TCF7</i>
OFF_T_4	3	aAggTCAGAGAAATTGGCCAAGG	chr5:173344963	-
OFF_T_5	3	GAgCTCAGAGAAATgGtCCAGGG	chr20:59763609	<i>PHACTR3</i>
OFF_T_6	3	GAaCTCAGAGgAATTGGCCtAGG	chr1:210923252	<i>KCNH1</i>
OFF_T_7	3	GATCTCAaAGAAATTGaCCcAGG	chr2:15921152	<i>MYCNUT</i>
OFF_T_8	3	GATCTgAGAGtcATTGGCCATGG	chr2:69202063	<i>ANTXR1</i>
OFF_T_9	3	GAaCTCAGAGAAgTgGGCCAGGG	chr12:2109540	<i>CACNA1C</i>
OFF_T_10	3	tATCTCAGAGAtcTTGGCCACGG	chr12:9329555	-
OFF_T_11	3	GAaCTCAGAGAAgTTGGCaATGG	chr12:48771787	<i>ADCY6</i>
OFF_T_12	3	aATCaCAGAGAAATTGGgCAGGG	chr12:120151784	<i>GCN1</i>
OFF_T_13	3	GAcCTCcGgGAAATTGGCCATGG	chr16:29977859	<i>TAOK2</i>
OFF_T_14	3	GATtTaAGAcAAATTGGCCAGGG	chr9:3091319	-
OFF_T_15	3	GATCcCAGAtAAATaGGCCAAGG	chr9:75480029	-
OFF_T_16	3	GATCaCAGgGAgATTGGCCATGG	chr11:5351762	<i>OR51B6</i>
OFF_T_17	3	cATaaCAGAGAAATTGGCCATGG	chr10:62596570	<i>LINC02929</i>
OFF_T_18	3	GAgCTCAagGAAATTGGCCAAGG	chr10:130480132	-

Table 38. Sequencing primers designed to study the off-target regions of the *TAOK1*_sgRNA1.

TARGET	SEQUENCE (5->3)	PRIMER	ANNEALING (°C)	AMPLICON LENGHT
-	AGCTGCAGTAGAAGAAAGCCT	OF_T_1_F	60	232
	TGTGAAGCCTGAACCAACCA	OF_T_1_R		
<i>UNC5D</i>	TGCCTATGCCACAGATTGCCT	OF_T_2_F	60	424
	CACGATGTCACAAACACGGT	OF_T_2_R		
<i>TCF7</i>	CGGCACCTACATCCCAGGT	OF_T_3_F	63	380
	AGATGCAGGCAGCCATAGGT	OF_T_3_R		
-	CCCATCCTCCTAGGTGTCCA	OF_T_4_F	60	485
	GGAAGGATGCCCCACATGAA	OF_T_4_R		
<i>PHACTR3</i>	AGTGCCTGGAACCCATGGTT	OF_T_5_F	60	999
	AGCTACTCAGCACTCCTCTGAC	OF_T_5_R		
<i>KCNH1</i>	CATGCCCAGTATTCTCCCC	OF_T_6_F	60	816
	GGACCCCCAATGGCAGTTTA	OF_T_6_R		
<i>MYCNUT</i>	TTGTTGTTCCGCCGTGTTTC	OF_T_7_F	60	700
	GCTTTGGTGCTCAAGCCTTC	OF_T_7_R		
<i>ANTXR1</i>	GTCAGGACCAGGGGGAAGCA	OF_T_8_F	63	401
	TGGCTTCCATGACCCCAGACT	OF_T_8_R		

CACNA1C	CCCTTTCCAATGCCAGGGGC	OF_T_9_F	63	517
	GCTGCAGTAGCCACCCTGAC	OF_T_9_R		
-	ACAACATGAAGTCCCAGCAAG	OF_T_10_F	60	236
	GGTTAATCCCAAGGCCAGCA	OF_T_10_R		
ADCY6	CTCCCCTGCTCCCCAACAGT	OF_T_11_F	63	505
	GGTGTCCACAGGCAACAGGG	OF_T_11_R		
GCN1	GCACCCACCACTAGATGTCCCA	OF_T_12_F	63	491
	GGTGTGAGAAAGCCTAAGCAGC	OF_T_12_R		
TAOK2	CAGCCCGCGTGATCTGGTTG	OF_T_13_F	63	519
	CCCATTGAGAGCCGGGCATG	OF_T_13_R		
-	ATTAGCCACATCTGCACGT	OF_T_14_F	60	936
	TGTGACTTCCAGGGAGAGGA	OF_T_14_R		
-	ACTCTCTCCATGCCTAGAATTT	OF_T_15_F	58	249
	ACTAAATATGACAGGCAAAGGGT	OF_T_15_R		
OR51B6	AGGCATGGAGAAGGCACATC	OF_T_16_F	60	639
	CCAGAACCAATGCCCATGACAG	OF_T_16_R		
LINC02929	TCTCGACCTAGGTGTCAGCA	OF_T_17_F	60	795
	AGGGGCACTGGAATATCCT	OF_T_17_R		
-	TGGATGGTTGTGAGCTCTTGGG	OF_T_18_F	60	730
	TCAGACAGGAGGGCGATACT	OF_T_18_R		

The absence of unexpected editing events in the 18 analysed off-target regions was confirmed through the results of Sanger sequencing (Figure 109).

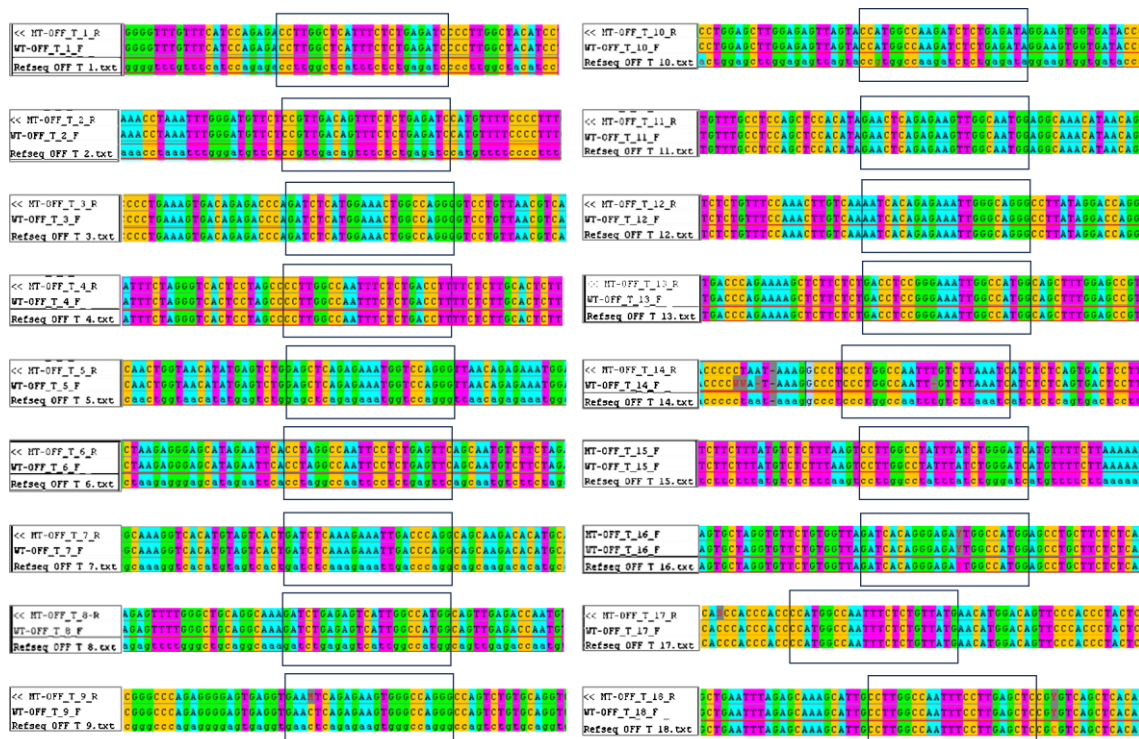


Figure 109. Analysis of the presence of editing events in off-target positions for the guide *TAOK1*_sgRNA1.

The region of each potential off-target was sequenced in the edited cell line Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1*-i2 (indicated as “MT” in the image), and on the isogenic control Ctrl.BS-PBiPS37-Sv4F-1 (indicated as “WT” in the image). Both sequences were aligned with the genomic reference sequence, previously indicated in Table 37. Black rectangles indicate the predicted off-target region. Image created by the author in Codon Code Aligner v10.0.2.

4.3.3. Evaluation of the functional consequences of *TAOK1* stop gain mutations in brain organoids.

The aim of this section is to comprehensively assess the functional consequences of stop gain mutations in *TAOK1* within human brain organoids. Integrating information from transcriptional and structural studies might uncover valuable insights into the role of *TAOK1* in normal brain development and its potential relevance to neurological conditions.

4.3.3.1. Assessment of alterations in brain organoid size.

The development of brain organoids was followed by taken representative images of the cultures on medium-change days. Throughout the initial 18 days of the experiment, spheroids were arranged in a 96-well plate, which allow to precisely track each one's development (Figure 110). Starting from day 18 onwards, growing organoids were placed in 6-well plates, and representative images of the each well were taken until they reached a size that exceeded the capabilities of the smaller microscope objective (Figure 111).

The existence of differences in the size of WT organoids and *TAOK1*-mutant organoids was quantified by measuring the area of the previously captured images, as shown in Figure 112. Data suggest that from day 18 until day 27, *TAOK1*-mutant organoids are larger than their isogenic control, but these differences are reduced overtime.

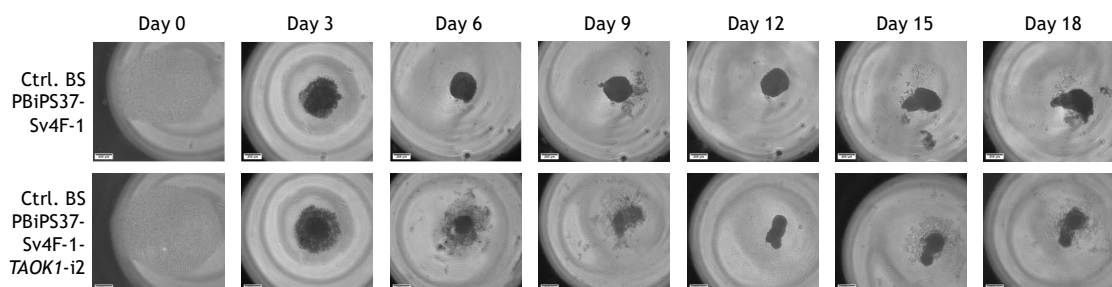


Figure 110. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and *TAOK1*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1*-i2) hiPSCs lines on the first 18 days of the protocol. The same well was imaged per cell line on each time point.

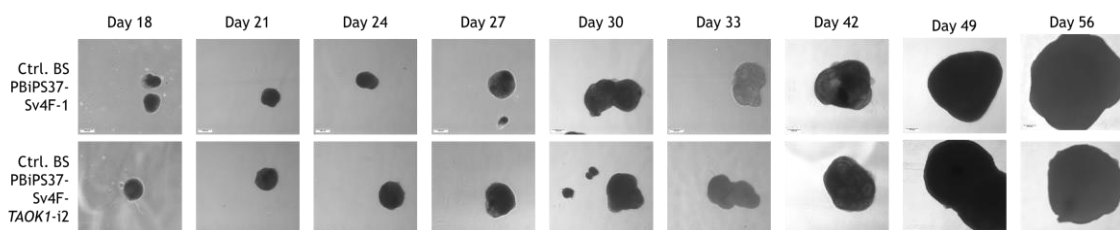


Figure 111. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and *TAOK1*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1*-i2) hiPSCs lines from day 18 to 56 of the protocol.

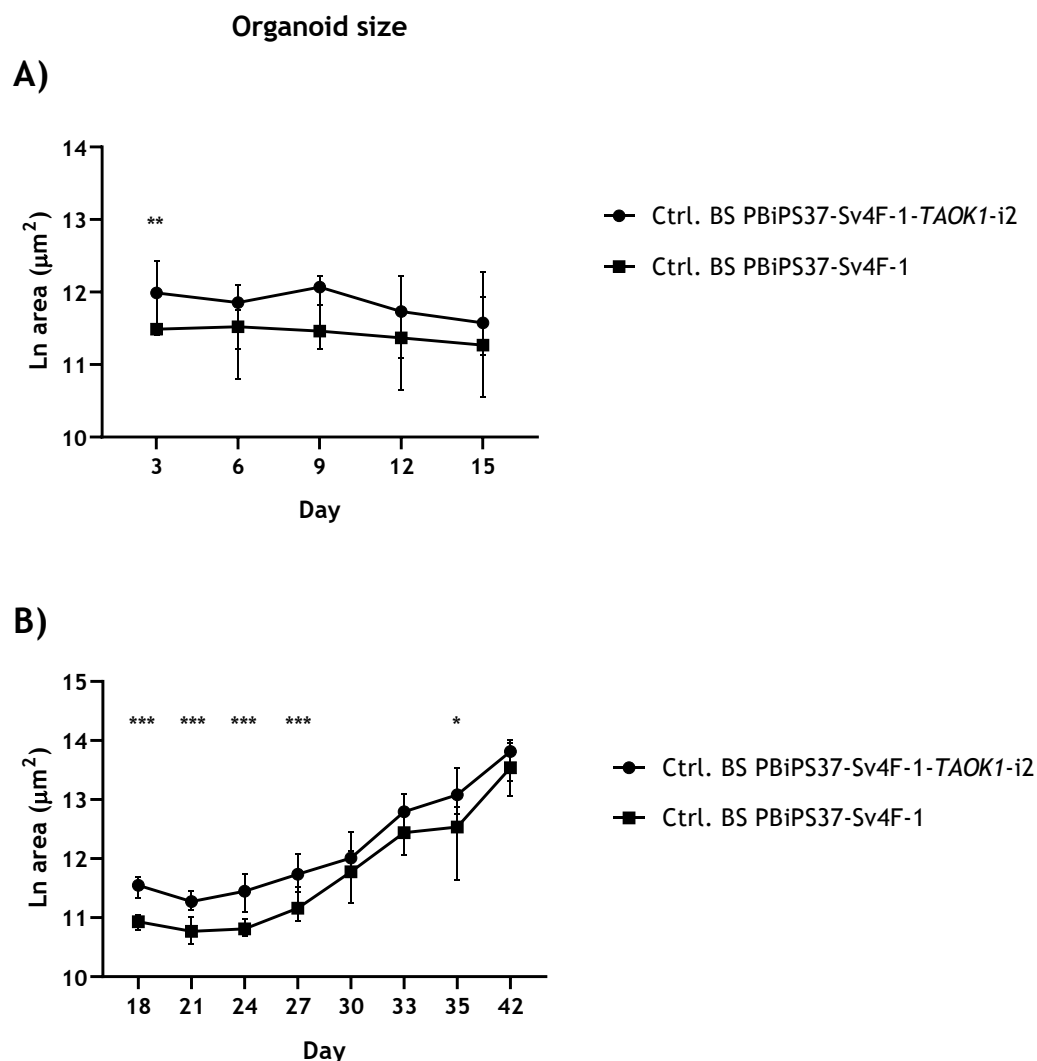


Figure 112. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and *TAOK1*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2*) hiPSCs lines. A) Days 3-15 of the protocol. B) Days 18-42 of the protocol. Area is represented in logarithm scale as data show a lognormal distribution. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

4.3.3.2. Immunohistochemistry characterization of human brain organoids with knockout mutations in *TAOK1*.

Following the visual assessment of developmental alterations *TAOK1*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2*) brain organoids, samples were taken to better characterize their internal structure and the presence of neuronal subtypes of interest.

First and foremost, the knockout of the *TAOK1* protein was verified in human brain organoids derived from the cell line Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2* (Figure 113).

To continue with the characterization, the presence of neural progenitor cells (SOX2, PAX6) and mature neurons (MAP2) was investigated in *TAOK1*-mutant brain organoids. The images presented in Figure 114 demonstrate that both types of cells are present in the WT organoids, as well as in *TAOK1*-mutants, with no significant differences at the analysed time

points (day 48 and day 62), represented in Figure 116. It is also important to note that both types of organoids present their $SOX2^+ PAX6^+$ neural progenitors arranged in similar ventricular-like structures, without apparent differences between them.

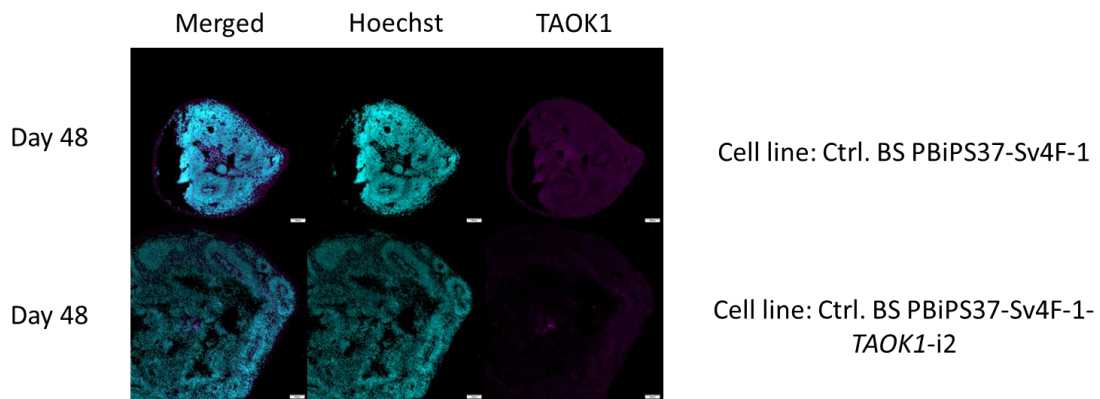


Figure 113. Immunofluorescence for the antigen TAOK1 in human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and edited (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1*-i2) cell lines.

The presence of GABAergic interneuron precursors ($DLX2^+$) and GABAergic neurons ($GAD67^+$) was also addressed in the generated 3D models (Figure 115, Figure 116). At day 48 of development, human brain organoids derived from *TAOK1*-mutant cell line show a significant reduction in the numbers of $DLX2^+$ and $GAD67^+$ cells. That difference is no longer apparent at day 62, suggesting a delayed but not impaired differentiation of GABAergic neurons in *TAOK1*-mutant brain organoids.

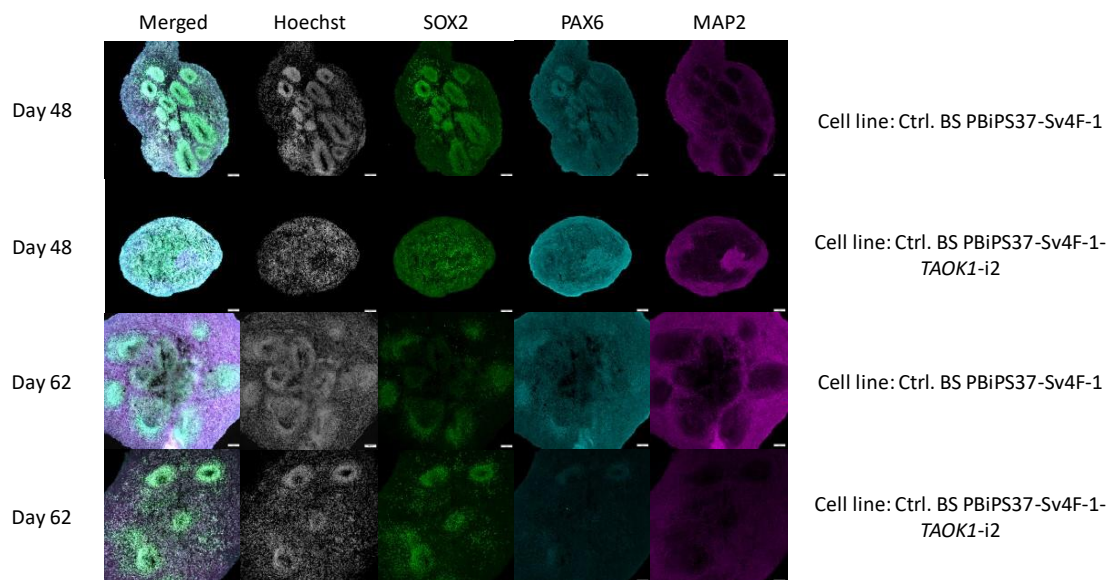


Figure 114. Characterization of the expression of cell-markers SOX2, PAX6, and MAP2 by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1*-i2) cell lines.

Finally, the expression of cortical-neuronal markers CTIP2/BCL11B and SATB2 was studied. As can be observed in Figure 117, both the WT and *TAOK1*-mutant brain organoids contain deep-layer ($CTIP2/BCL11B^+$) and upper-layer ($SATB2^+$) neurons, located in the periphery of the progenitor-enriched rosette-like structures.

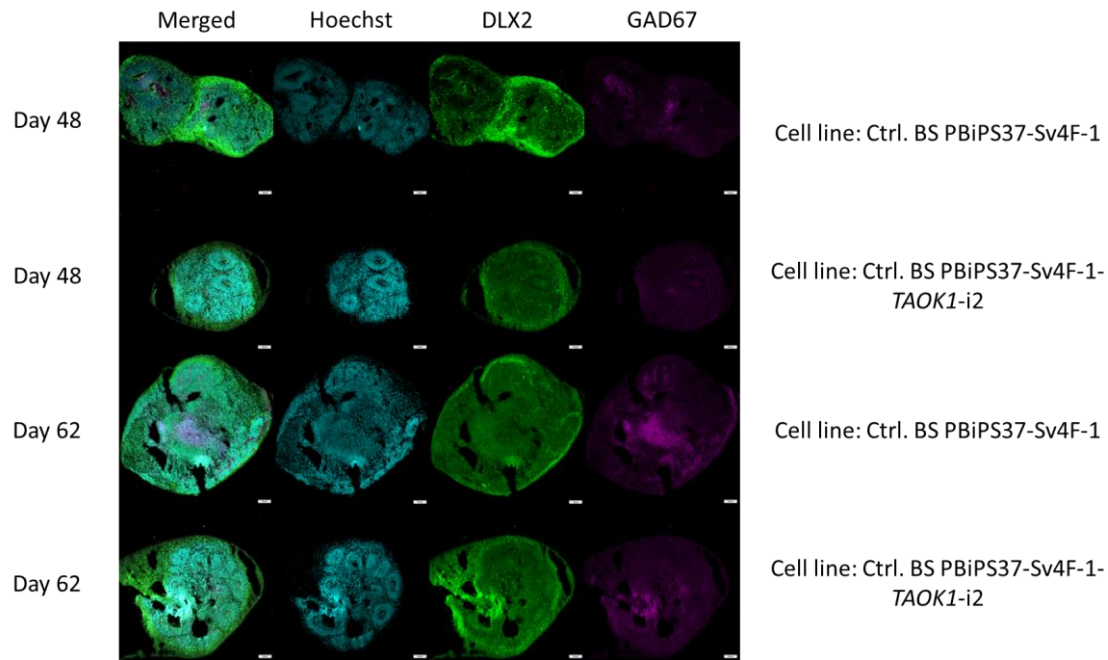


Figure 115. Characterization of the expression of cell-markers DLX2 and GAD67 by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2) cell lines.

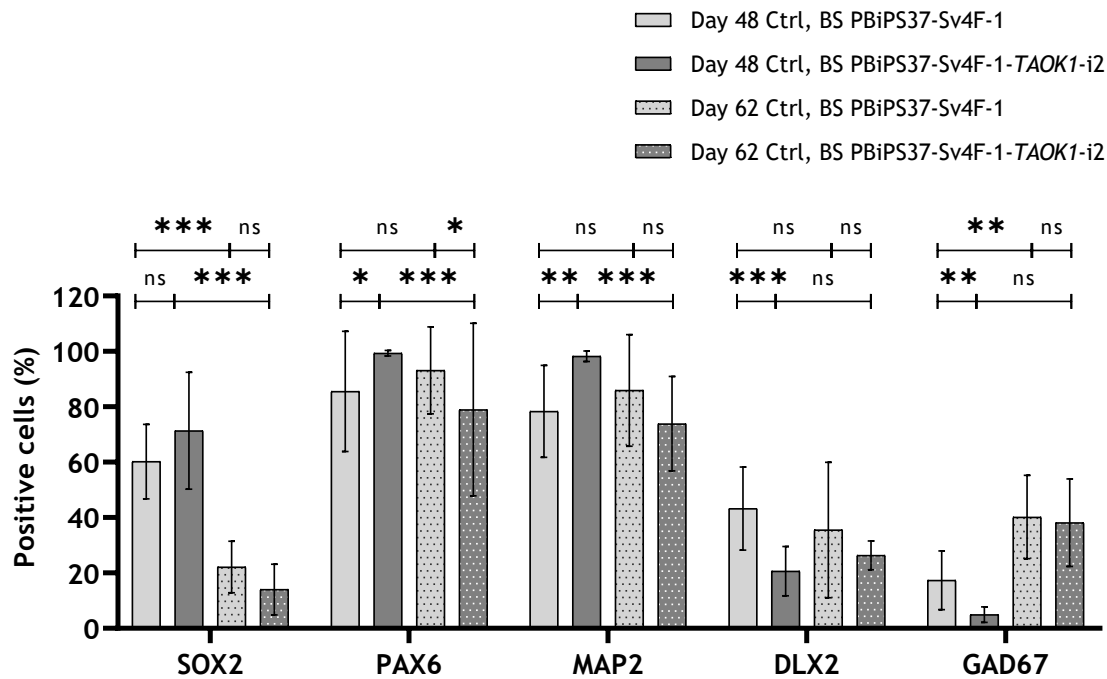


Figure 116. Quantification of the expression the cell-markers SOX2, PAX6, and MAP2 in human brain organoids. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2) cell lines. Data represented as mean $\pm$ SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

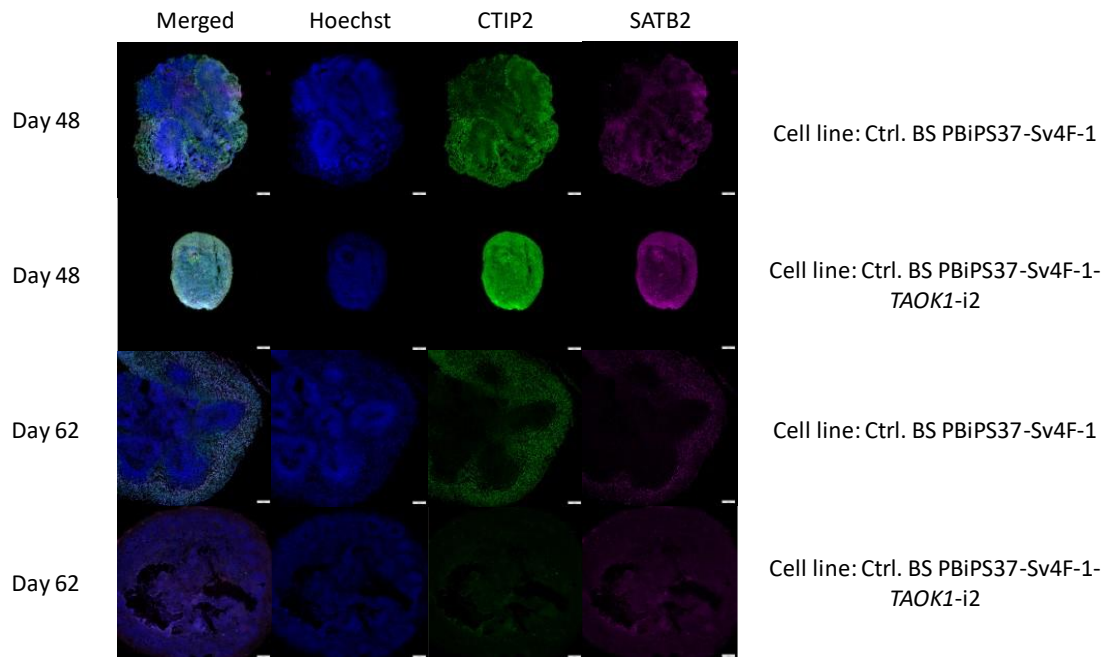


Figure 117. Characterization of the expression of cell-markers CTIP2/BCL11B and SATB2 by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2*) cell lines.

4.3.3.3. Transcriptomic analysis of *TAOK1*-mutant brain organoids by means of scRNA-seq.

Transcriptomic landscape and cell-type composition of human brain organoids derived from the Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2* line and its isogenic control (Ctrl.BS-PBiPS37-Sv4F-1) were analyzed using scRNA-seq at two time points of interest: day 48 and day 62.

Sequencing data was processed in the *Universidad de Málaga* by James Perkins, PhD. and Sergio Alías, PhD. After quality control, samples were integrated based on the cell-line or time point of interest and represented using Uniform Manifold Approximation and Projection (UMAP). Then, in our laboratory, cell-types were assigned to each cluster based on their top 10 differentially expressed genes^{292–296}, along with canonical cell-type markers^{119,297}.

In this section, the results of scRNA-seq data analysis will be presented, commencing with cell-line integrated data analysis, and proceeding to integrated data analysis across different time points. Unfortunately, these results arrived just two months before this thesis needed to be deposited, so only a preliminary analysis of the data could be included.

Cell-line integrated data analysis was independently performed for Ctrl.BS-PBiPs37-Sv4F-1 and Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2* samples, as illustrated in Figure 118. First, cell-type composition of Ctrl.BS-PBiPs37-Sv4F-1 brain organoids will be commented, followed by Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2*.

Based on the analysis of differentially expressed genes, cells originating from Ctrl.BS-PBiPs37-Sv4F-1 brain organoids were categorized into 23 clusters, including subcortical and cortical cell populations (Figure 118, Figure 119). Notably, distinct clusters (3, 4, 6, 10, 16, 18, 22, 23) were identified within subcortical populations, demonstrating a clear separation from cortical cell types. Among these, clusters 6, 10, 22, and 23 exhibited expression of retina markers (*EYS*, *EGFLAM*, *PDC*, *IMPG2*). Despite their apparent proximity in the UMAP plots, these clusters differ in their cell composition. Specifically, cluster 6 was enriched in day 48 cells, while clusters 10 and 22 showed enrichment in day 62 cells.

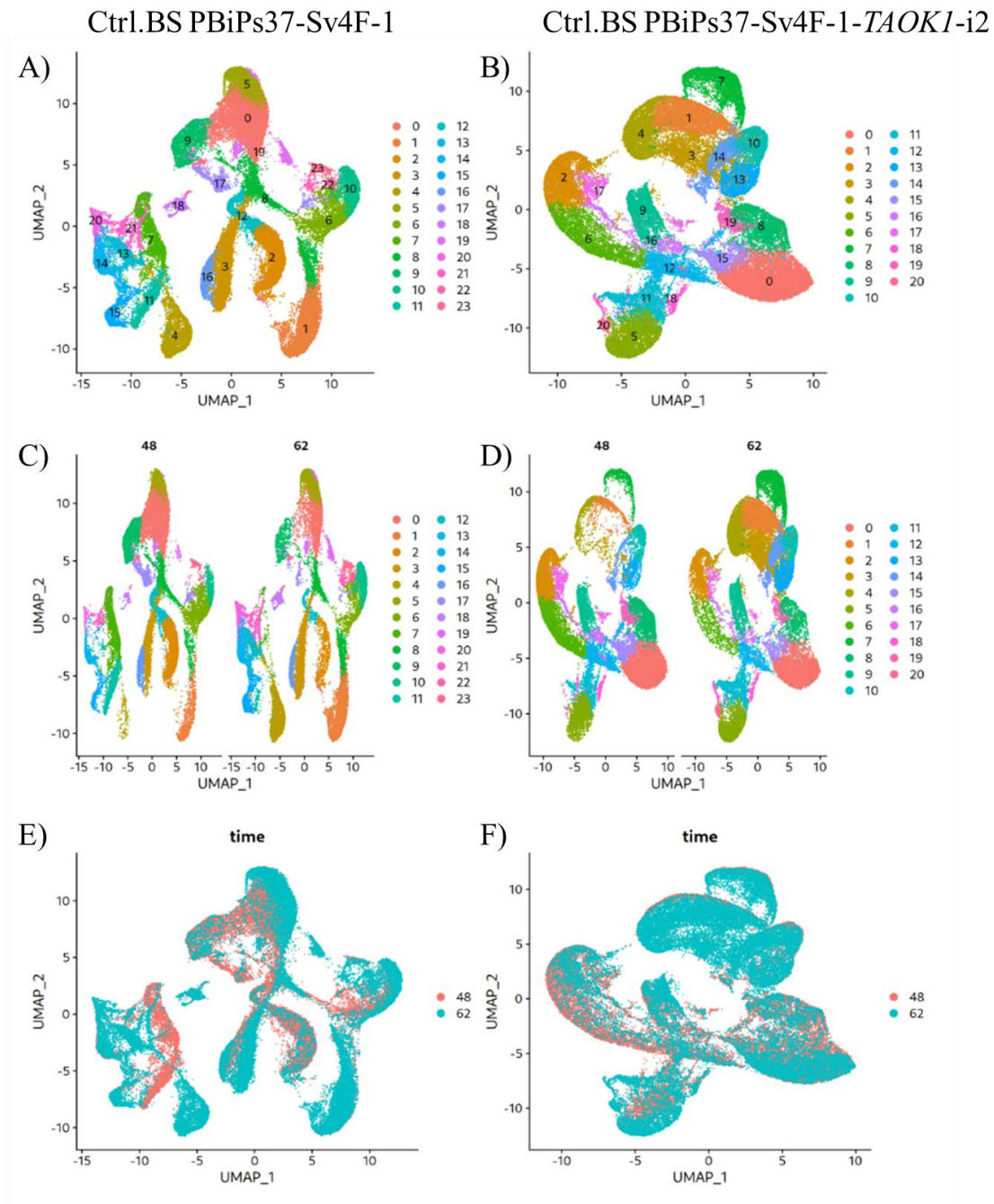


Figure 118. UMAP representations of scRNA-seq data integrated by cell line: Ctrl.BS-PBiPs37-Sv4F-1 (A, C, E) and Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2 (B, D, F). A-B) Clusters identified on each group of samples with integrated information of both time point (day 48, day 62). C-D) Clusters identified on each group of samples with separated representations for both time points (day 48, day 62). E-F) Day 48 and day 62 distribution in the UMAP.

Retinal cells were not the only subcortical cell population found in Ctrl.BS-PBiPs37-Sv4F-1 samples. Notably, clusters 3, 4 and 16 show gene expression signatures compatible with subcortical interneurons (*TFAP2A*, *GAD2*). Clusters 3 and 4 are enriched in cells obtained from day 62 organoids, while in cluster 16 are more abundant day 48 cells.

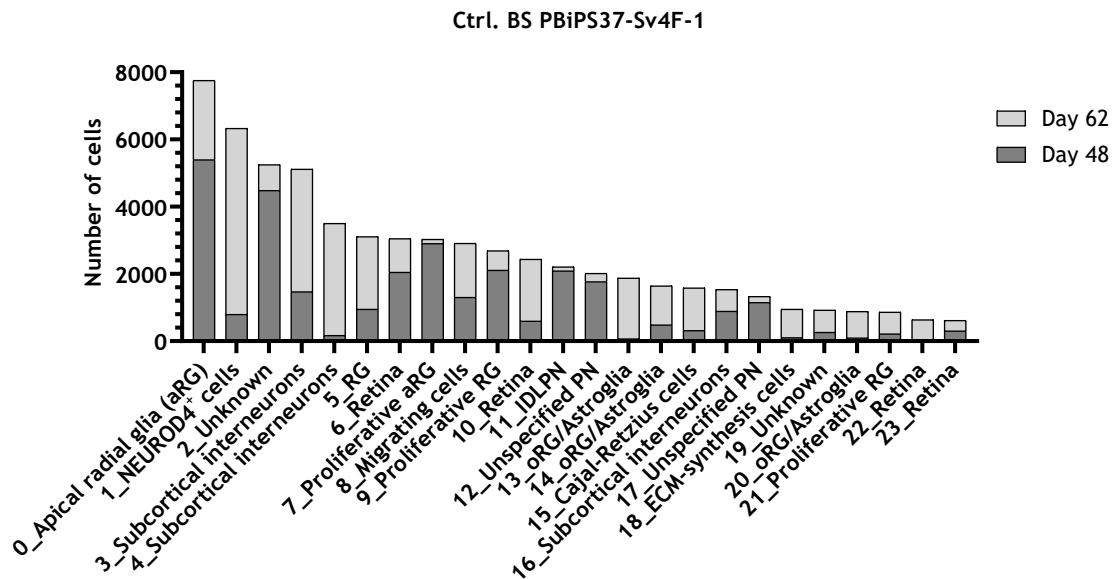


Figure 119. Clusters identified after integrating scRNA-seq Ctrl.BS-PBiPS37-Sv4F-1 samples. The number of cells from each time point analysed (day 48, day 62) is represented for each cluster. Day 48 samples: dark grey; day 62 samples: light grey.

Cortical subpopulations identified in Ctrl.BS-PBiPs37-Sv4F-1 samples include cells in different maturation stages: from neural stem cell populations (radial glia, RG) to projection neurons (PN).

Five clusters of RG cells were identified in Ctrl.BS-PBiPs37-Sv4F-1 samples (0, 5, 7, 9, 13, 14, 20, 21). Two of them, cluster 0 and 7, were characterized by the expression of apical radial glia (aRG) cell-type markers (*SOX2*, *HES1*). In both cases, day 48 derived cells enriched both clusters, with was expected since aRG cells are more abundant on the early stages of brain organoid development. In clusters 5, 9 and 21, aRG markers were identified along with outer radial glia (oRG) cell-type markers (*HOPX*, *PEA15*, *MOXD1*, *LGALS3BP*), thus classifying the population as RG. Clusters 5 and 21 are enriched in day 62 cells, while cluster 9 is mainly formed by day 48 derived cells. Additionally, clusters 7, 9 and 21, also expressed proliferation markers (*MKI67*, *TOP2A*). Three clusters (13, 14, 20) were characterized by the expression of both oRG and astroglia markers (*GJA1*, *S100B*, *SPARC*), all of them enriched in day 62 cells, which is consistent with the maturation of brain organoids.

Two groups of unspecified PN were identified (12, 17), which were characterized by the expression of general neuronal markers (*EBF1*, *NEFL*, *CNTN2*, *SNCG*, *PRPH*). Both clusters are mainly composed of day 48 derived cells. On the other hand, cells from cluster 11 expressed immature deep layer projection neuron (IDLPN) markers (*PDE1A*, *NEUROD2*, *NEUROD6*, *TBR1*), along with *CTIP2/BCL11B* expression, which could indicate commitment to the deep-layer corticofugal projection neuron (CFuPN) lineage. This cluster was also enriched in day 48 cells. Cells from cluster 15 were mainly derived from day 62 brain organoids and were characterized by a strong expression of *RELN*, thus being classified as Cajal-Retzius cells.

Cluster 8 cells were characterized by the expression of *GADD45G*, which could be an indicative of neuronal migration. Cluster 1 cells were characterized by the expression of *NEUROD4* (expressed by neural progenitors) but lack other canonical cell subtype markers.

In addition, cluster 18 cells (enriched in day 62 cells) were characterized by the expression of genes involved in extracellular matrix (ECM) synthesis (*MGP*, *COL1A1*, *COL3A1*, *COL6A3*, *COL1A2*). Notably, this cluster is isolated from the others in the UMAP representation.

The remaining two clusters (2 and 19) could be matched to any subpopulation yet, and thus were identified as unknown in Figure 119.

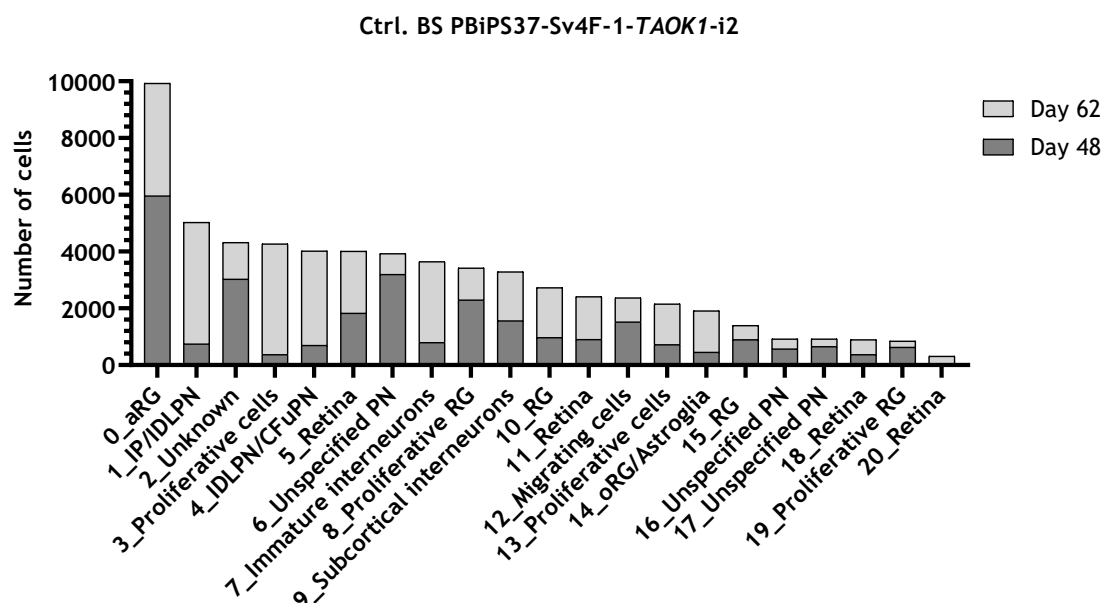


Figure 120. Clusters identified after integrating scRNA-seq Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2 samples. The number of cells from each time point analysed (day 48, day 62) is represented for each cluster. Day 48 samples: dark grey; day 62 samples: light grey.

Cells derived from Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2 brain organoids were grouped into 20 clusters (Figure 118, Figure 120), including subcortical and cortical cell types.

As it happened previously with cells derived from WT brain organoids, four clusters (5, 11, 18, and 20) were characterized by the expression of retina markers (*EYS*, *EGFLAM*, *PDC*, *IMPG2*). Cluster 20 is notably enriched in cells derived from day 62 organoids, while the remaining retina clusters have more equilibrated composition. Subcortical interneurons were also identified in Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2 brain organoids, as cluster 9 was characterized by the expression of *TFAP2A*. Noticeably, *GAD2* expression was very low in this cluster.

In Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2 brain organoids, six clusters of RG cells were identified (0, 8, 10, 14, 15, 19). Among those, cluster 0 was characterized by the expression of aRG cell-type markers (*SOX2*, *HES1*) and was enriched in day 48 derived cells, as expected. Clusters 8, 10, 15, and 19 expressed markers of both aRG and oRG (*HOPX*, *PEA15*, *MOXD1*, *LGALS3BP*) cell subtypes, so they were classified as RG. Interestingly, cluster 10 is enriched in day 62 derived cells, while cluster 15 is mainly formed by cells from day 48, which could be indicative of intermediate states between aRG and oRG. Day 48 cells are also the main component of both cluster 8 and 19, which also show a high expression of genes involved in cell division (*MKI67*, *TOP2A*), thus indicating the proliferative state of those RG populations. Finally, cluster 14 was enriched in day 62 derived cells characterized by the expression of oRG and astroglia (*GJAI*, *S100B*, *SPARC*) markers.

In addition, three clusters (6, 16, and 17) of unspecified PN enriched in day 48 cells were identified based on the expression of general neuronal markers (*EBF1*, *NEFL*, *CNTN2*, *SNCG*, *PRPH*), but lacking subtype identifiers. On the other hand, cells from cluster 1 expressed markers of intermediate progenitors (IP: *EOMES*, *INSMI*, *EMX1*) and IDLPN (*PDE1A*,

NEUROD2, *NEUROD6*, *TBR1*). Cells from cluster 4 also expressed IDLPN markers, in combination with a strong *CTIP2/BCL11B* expression, indicating commitment to the CFuPN lineage. Both clusters (1 and 4) are mainly composed of cells derived from day 62 Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2* brain organoids. Cluster 12 cells were characterized by the expression of *GADD45G*, which could be an indicative of neuronal migration.

Based on the top 10 differentially expressed genes and canonical markers, it was not possible to identify the cell class for clusters 2, 3, and 13, although the latter two expressed genes involved in cell proliferation.

As previously mentioned, scRNA-seq data was also integrated according to the time-points of interest. Time-point integrated data analysis was independently performed for day 48 and day 62 brain organoids, as represented in Figure 121. First, cell-type composition of day 48 brain organoids will be commented, followed by day 62, focusing on differences between Ctrl.BS-PBiPS37-SvF4-1 and Ctrl.BS PBiPS37-SvF4-1-*TAOK1-i2* brain organoids.

Cells derived from day 48 brain organoids were grouped into 22 clusters of cortical and subcortical cell populations (Figure 121, Figure 122). Of those clusters, four (5, 13, 21, and 22) were characterized by the expression of retina markers (*EYS*, *EGFLAM*, *PDC*, *IMPG2*). Cells originated in WT and *TAOK1*-mutant organoids are present in the four retina clusters, although the first are more abundant. Another cluster of subcortical neurons was identified in day 48 samples: cluster 2. Cluster 2 is composed by cells characterized by the expression of *TFAP2A*, a marker of subcortical cell identity. This cluster is slightly enriched in cells derived from WT brain organoids.

Eight clusters of RG cells were identified (0, 3, 4, 10, 11, 16, 17, 18) in samples obtained from day 48 brain organoids. Two of them, clusters 0 and 10 were enriched in WT-derived cells and expressed markers of aRG (*SOX2*, *HES1*). Five clusters (3, 4, 11, 16, 18) expressed markers of both aRG and oRG (*HOPX*, *PEA15*, *MOXD1*, *LGALS3BP*), and thus were classified as RG. Moreover, clusters 3, 11 and 16 expressed proliferation markers (*MKI67*, *TOP2A*). Except for cluster 11, *TAOK1*-mutant cells are the main component of all RG clusters. Cluster 17 cells expressed oRG and astroglia (*GJAI*, *S100B*, *SPARC*) markers and, interestingly, is mainly composed of WT cells.

In addition, three clusters (6, 12, and 14) of unspecified PN were found in day 48 samples, which expressed general neuronal markers (*EBF1*, *NEFL*, *CNTN2*, *SNCG*, *PRPH*). Clusters 6 and 14 were enriched in *TAOK1*-mutant cells, while WT cells were more abundant in cluster 12. The two clusters (8, 19) expressing IDLPN signature markers (*PDE1A*, *NEUROD2*, *NEUROD6*, *TBR1*) were also enriched in WT cells. Interestingly, cluster 15 was characterized by the expression of immature interneuron markers (*DLX2*, *DLX5*, *GAD2*) and was mainly formed by *TAOK1*-mutant cells.

Cluster 20 cells were characterized by the expression of *NEUROD4* but lack other canonical cell subtype markers. This cluster show the major difference in composition between *TAOK1*-mutant (8.1 %) and WT cells (91.9 %). Cluster 9 was characterized by the expression of *GADD45G*, which could be an indicative of neuronal migration, and showed a quite equilibrated composition of WT and *TAOK1*-mutant cells. Unfortunately, clusters 1 and 7 couldn't be identified yet based on the top 10 differentially expressed genes and canonical markers.

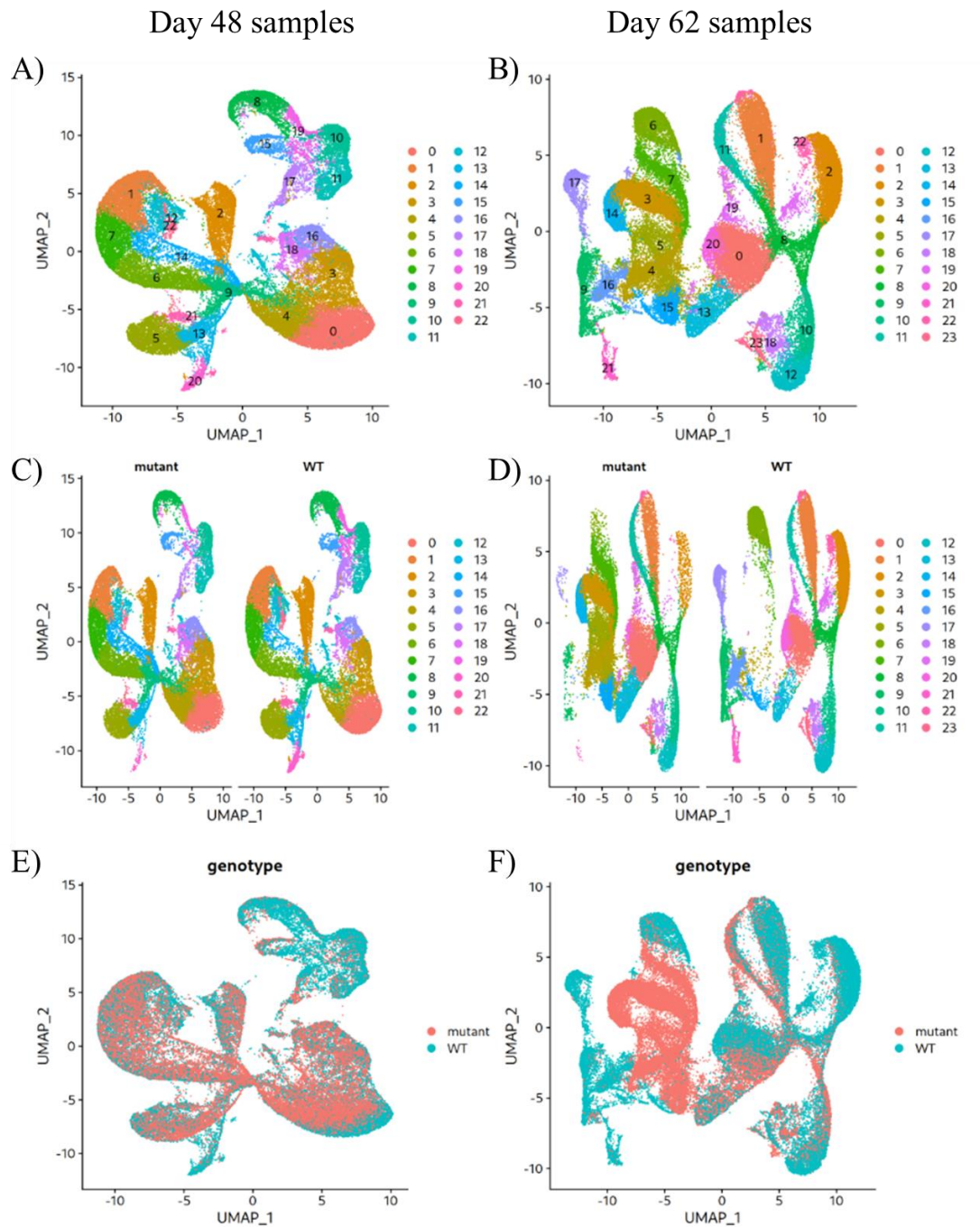


Figure 121. UMAP representation of scRNA-seq data integrated by time: day 48 samples (A, C, E) and day 62 samples (B, D, F). A-B) Clusters identified on each group of samples with integrated information of both cell lines (Ctrl.BS-PBiPs37-Sv4F-1, Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2). C-D) Clusters identified on each group of samples with separated representations for both cell lines (Ctrl.BS-PBiPs37-Sv4F-1 identified as “WT”, Ctrl.BS-PBiPs37-Sv4F-1-TAOK1-i2 identified as “mutant”). E-F) Cell line distribution in the UMAP.

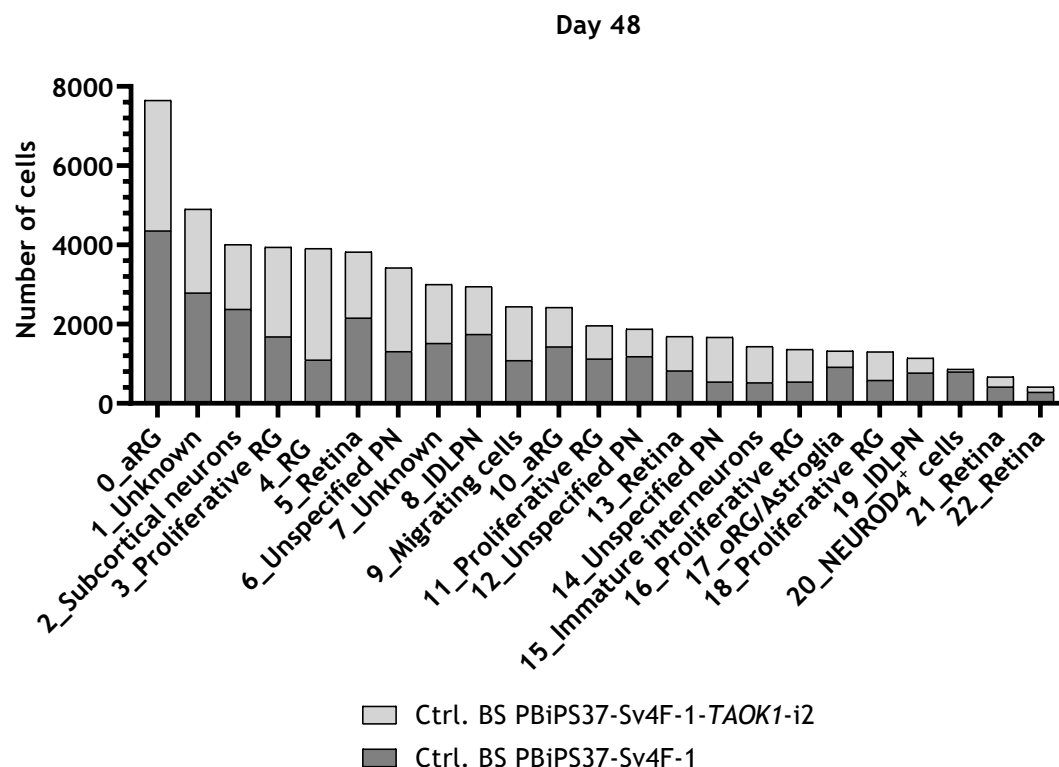


Figure 122. Clusters identified after integrating scRNA-seq day 48 samples. The number of cells from each cell line (Ctrl.BS-PBiPS37-SvF4-1, Ctrl.BS-PBiPS37-SvF4-1-TAOK1-i2) is represented for each cluster. Ctrl.BS-PBiPS37-SvF4-1 samples: dark grey; Ctrl.BS-PBiPS37-SvF4-1-TAOK1-i2 samples: light grey.

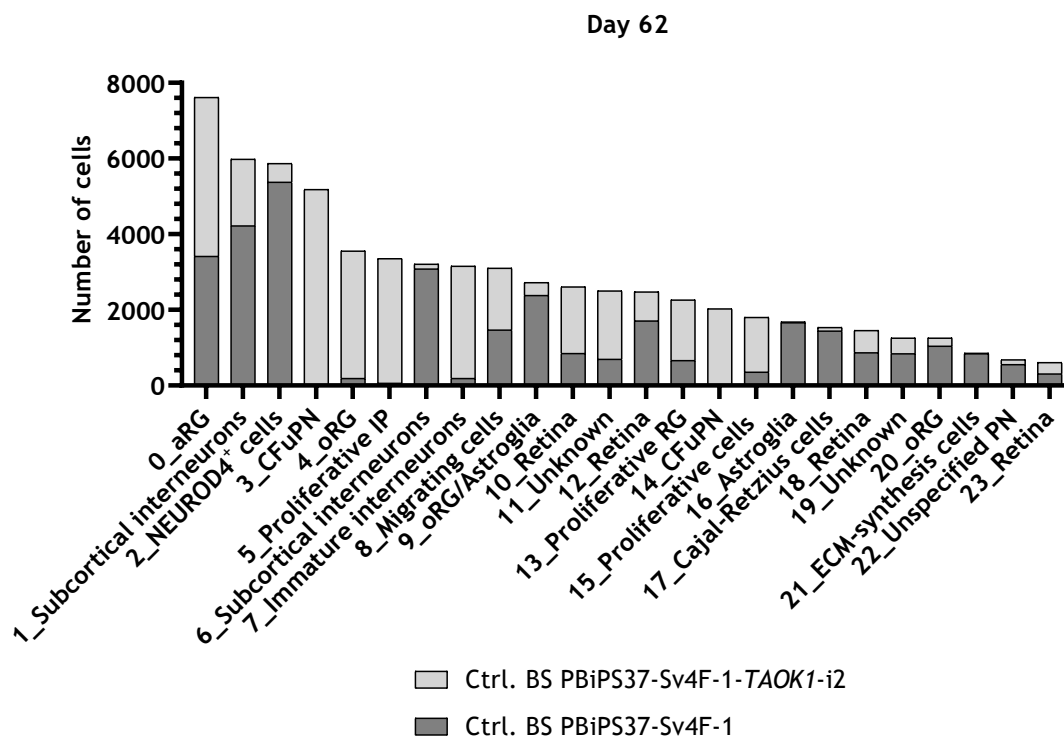


Figure 123. Clusters identified after integrating scRNA-seq day 62 samples. The number of cells from each cell line (Ctrl.BS-PBiPS37-SvF4-1, Ctrl.BS-PBiPS37-SvF4-1-TAOK1-i2) is represented for each cluster. Ctrl.BS-PBiPS37-SvF4-1 samples: dark grey; Ctrl.BS-PBiPS37-SvF4-1-TAOK1-i2 samples: light grey.

Finally, cluster composition of day 62 brain organoids will be addressed. As it happened with day 48 samples, day 62 cells are divided into 22 clusters (Figure 121 and Figure 123). In terms of subcortical populations, four clusters (10, 12, 18, 23) were identified as retina cells due to the expression of subtype specific markers, such as *EYS*, *EGFLAM*, *PDC*, and *IMPG2*. In addition, two subcortical interneuron clusters (*TFAP2A*, *GAD2*) were identified, cluster 1 and 6, both mainly composed of cells derived from Ctrl.BS-PBiPS37-SvF4-1 brain organoids.

In day 62 samples, 6 clusters show expression patterns characteristic of RG cells (0, 4, 9, 13, 16, 20). On one side, cluster 0, which show a major proportion of *TAOK1*-mutant cells, express genes associated with aRG cell identity (*SOX2*, *HES1*). On the other side, clusters 4 and 20 were identified as oRG (*HOPX*, *PEA15*, *MOXD1*, *LGALS3BP*). Interestingly, *TAOK1*-mutant cells are mainly located in cluster 4, while WT cells mainly form cluster 20. Proliferative associated genes (*MKI67*, *TOP2A*) were identified along with RG markers in cluster 13, formed mainly by *TAOK1*-mutant cells. Additionally, astroglia markers (*GJA1*, *S100B*, *SPARC*) were found in clusters 9 and 16, both formed mainly by WT cells.

Only one cluster of unspecified PN was identified: cluster 22. The cells from this cluster had mainly a WT origin and expressed general neuronal markers (*EBF1*, *NEFL*, *CNTN2*, *SNCG*, *PRPH*). On the contrary, clusters 3, 5 and 14 were enriched in *TAOK1*-mutant cells, being WT cells absent from the first cluster. Cluster 5 cells displayed a gene expression pattern coherent with proliferative IP cells (*EOMES*, *INSMI*, *EMX1*, *MKI67*, *TOP2A*), while clusters 3 and 14 expressed deep-layer CFuPN markers (*BCL11B*, *SOX2*, *LDB2*).

Cells from cluster 7 expressed immature interneuron markers (*DLX2*, *DLX5*, *GAD2*). This subpopulation was mainly formed by *TAOK1*-mutant cells, which is consistent with findings from day 48 samples.

Clusters 2, 17, and 21 were enriched in WT cells. Cells from cluster 2 were characterized by the expression of *NEUROD4*, which is associated with neural progenitors, while cluster 17 mainly expressed *RELN*, thus being classified as Cajal-Retzius cells. Cluster 21 was characterized by the expression of genes involved in ECM synthesis (*MGP*, *COL1A1*, *COL3A1*, *COL6A3*, *COL1A2*). Cells from cluster 8 expressed *GADD45G*, which could be an indicative of neuronal migration. This cluster was formed by WT and *TAOK1*-mutant cells in a similar proportion.

The information used to classify the aforementioned clusters was not enough to identify the cell subtype identity of clusters 11, 19, and 15, although the latter expressed proliferation markers.

4.3.4. Generation of an *in vivo* zebrafish model to study *taok1*.

The present subheading is dedicated to the generation of a mutant zebrafish model to characterize the function of *taok1* (*TAOK1*'s orthologous gene) and its role during neurodevelopment.

4.3.4.1. Evaluation of the presence of *TAOK1* orthologous sequences in zebrafish.

The human gene *TAOK1* has one orthologous high confidence gene in zebrafish, known as *taok1*. As the public database ENSEMBL²⁰⁶ indicates, 84.92 % of the human gene sequence matches its corresponding orthologous in zebrafish, with the homology being also high at both the cDNA (72 %) and protein (84 %) levels. The alignment between the human and zebrafish proteins is represented in Figure 124, on which it can be observed the high homology of both proteins, including the protein kinase domain marked in orange.

```

sp|Q7L7X3|TAOK1_HUMAN      MPSTNRAGSLKDPETAELEFKEDPEKLTDLREIGHGSFGAVYFARDVRTNEVVAIKKMS
tr|X1WGA8|X1WGA8_DANRE    MPNAVRAGSLKDPDIADLFKEDPERLFSDLREIGHGSFGAVYFARDVRTSEVVAIKKMS
**:*:*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      YSGKQSTEKWDIIEVKFLQRIKHPNSIEYKGCYLREHTAWLVMCYLGSASDLLEVHK
tr|X1WGA8|X1WGA8_DANRE    YSGKQSEKWDIIEVKFLQRIKHPNSIEYKGCYLREHTAWLVMCYLGSASDLLEVHK
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      KPLQVEITAAITHGALQGLAYLHSHMTIHRDIKAGNILLTEPGQVKLADFGSASMASPAN
tr|X1WGA8|X1WGA8_DANRE    KPLQEIETAAITHGALQGLAYLHSHMTIHRDIKAGNILLTEPGQVKLADFGSASIASPAN
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      SFVGTPYMAPEVILAMDEGQYDGKVDVWSLGTICIELAERKPLFMNMMSALYHIAQN
tr|X1WGA8|X1WGA8_DANRE    SFVGTPYMAPEVILAMDEGQYDGKVDIWSLGTISIELAERKPLFMNMMSALYHIAQN
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      ESPTLQSNVSDYFRNFVDSCLQKIPQDRPTSEELKHIFVLRERPETVLIDLIRTKDA
tr|X1WGA8|X1WGA8_DANRE    ESPSLQSEWTDYFRNFVDSCLQKLPQDRPHEELKHAFVLRERPESVLIDLIRTKDA
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      VRELDNLQYRKMKLLFQEAHNGPAVEAQEEEEEQDHGVRTGTVNSVGSNQSIPTMSIS
tr|X1WGA8|X1WGA8_DANRE    VRELDNLQSRKMKLLFQEAHNGPAVEVPDDEEEVHSGRTGTVNSVGSNQSIPTMSIS
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      ASSQSSVNSLPOVS-----DQKSELDMEGHTVMSNSVIHLKPEEENYREGDPRTRA
tr|X1WGA8|X1WGA8_DANRE    ASSQSSVNSLPDADTDQNKSEVDLMEKDTIIPNSSIHWKPEEESYQEEETEVQKP
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      SDPQSPQVSRHKSHYRNREHFATIRASLVTRQIHEQDSLEQMSGKRMRRQHQK
tr|X1WGA8|X1WGA8_DANRE    TEPQSPPAHT-PRQYRNREHFATIRASLVTRQIHEQDSVREQMSGKRMRRQHQK
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      QLMTLENLKAEMDEHRLRLDKDLTQRNFAAEMEKLIKQQAAMEKAQVMSIEEKKF
tr|X1WGA8|X1WGA8_DANRE    HLLALENLKGMDEHRLRLDKELNQRNFAAEMDKLIKQGTMEKDAKTFANDEKFF
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      QQHIQAQKKELNSFLESQKREYKLRKEQLKEELNENQSTPKCKEQEHLKQKELIQHFQ
tr|X1WGA8|X1WGA8_DANRE    LQHIQAQKKELNSFLESQKREYKLRKEQLKEELNENQSTPKCKEQEHLKQKELIQHFQ
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      AEEAANLRRQRYLEECRRFKRRLGRHNLQDLVREELNKRQTKQDLHAMLLRQH
tr|X1WGA8|X1WGA8_DANRE    AEEAANLRRQRYLEECRRFKRRLIARHNVQDLVREELNKRQTKQDLHAMLLRHH
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      ESMQLEFRHLNITQKMRCELIRLHQTELTNQLYENKRERELRRKHVMEVRQPKSLK
tr|X1WGA8|X1WGA8_DANRE    ESMQLEFRHLITQKMRQELIRLHQTELTNQLYENKRERELRRKHVMEVRQPKSLK
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      SKELQIKKQFQDCKIQTRQYKALRNHLLSTPKSEHKAVLKRLEEQTRKLALAEQYD
tr|X1WGA8|X1WGA8_DANRE    SKELQIKKQFQDACKIQTRQYKALRNHLLSTPKPEHKTMLKRLKEEQTRKLALAEQYD
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      HSNEMLSQALRLDEAQAECCQLKMLQQLLELLNAYQSKIKMQAEAQHRELRELEQ
tr|X1WGA8|X1WGA8_DANRE    HTINMLSQALRLDEAQAECCQLRMQLQQLLELLNAYQSKIKMQTDAQERERLELEQ
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      RVSLRRALLEQTEEMALQNERTERIRSLERQARETEAFDSESMRLGFSNMVLNLS
tr|X1WGA8|X1WGA8_DANRE    RVSLRRALLEKTEEMLTQNERTERIRSLERQARETEAFDSESMRLGFSNMVLNLS
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      PEAFSHYPGASG-W---SHUPTGGPGHGHMGGP-----PQAWGHPMQ
tr|X1WGA8|X1WGA8_DANRE    PEGLSHSPGAPASWPHQHSGSGTQSGSTGSGGHHYHPSQGAQVQAWGQAIP
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      G-GPQWGHPSGPMQGVPRGSMGVNRSQALRRITASGRTEQMSRSTSVTSQISNGSH
tr|X1WGA8|X1WGA8_DANRE    GGLPPWGHSSGPPGAMPKSSVMNRNSQALRRITASGRTEQMSRSTSVTSQISNGSH
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

sp|Q7L7X3|TAOK1_HUMAN      MSYT      1001
tr|X1WGA8|X1WGA8_DANRE    LSYT      1023
*****

```

Figure 124. Alignment of the proteins TAOK1 (above) and taok1 (below) generated by the Clustal Omega program (EMBL-EBI) ³⁰¹. In orange it is marked the protein kinase domain of the protein (28 - 281 aa).

4.3.4.2. CRISPR-Cas9 guide design and synthesis.

Once verified the high homology between *TAOK1* and its orthologous sequence *taok1*, it was necessary to design CRISPR-Cas9 guides to knockout the gene in zebrafish.

Guides were initially designed with the CRISPRscan webtool²⁹⁸, which identified a total amount of 220 potential guides to target *taok1*. To select the most fitting guides for the proposed experimental approach, several filters were applied. First, only guides that target simultaneously the three transcripts of *taok1* were considered as possible candidates. The next applied filter

discarded guides with predicted off-targets in the zebrafish genome. Finally, the remaining guides were ordered based on their predicted efficiency, based on their CRISPRscan score (between 0 - 100). After selecting the most efficient guides (CRISPRscan score > 70), 32 guides remained on the list.

To further help with the selection process, those guides were analysed by the Synthego's Verify Guide Designing webtool²⁸² with the objective of corroborating their suitability for the knockout generation. Using this webtool 10 guides were discarded since they did not fit the filtering criteria (target all transcripts, high efficiency, and minimal off-targets) with the new algorithm.

As a last step in the guide selection process, guides were filtered considering their target coordinates. Since the function of *taok1* relies on the protein kinase activity (84 – 297 aa of *taok1*), two guides that target the coding region of that domain were selected (exons 1 – 10). In Figure 125 can be found a schematic representation of the CRISPR-Cas9 machinery, along with the selected target regions in the zebrafish genome for the selected guides (*taok1*_sgRNA1 and *taok1*_sgRNA2).

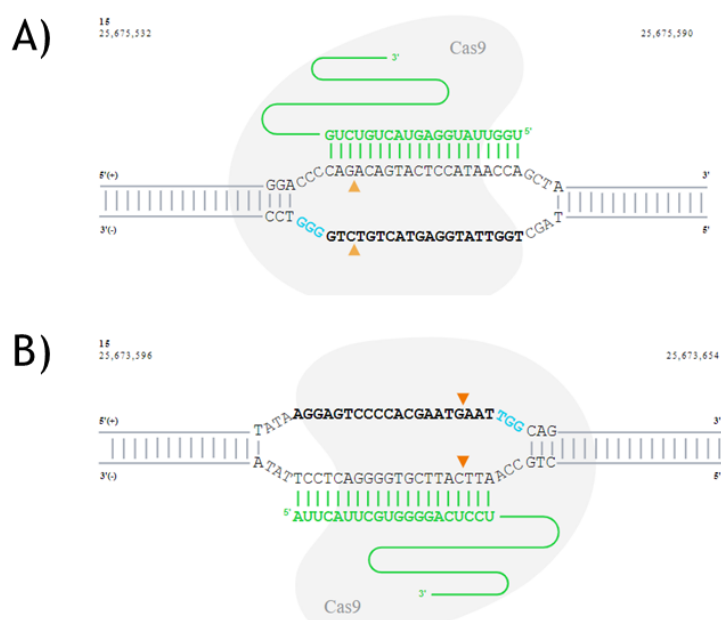


Figure 125. Schematic representation of the sgRNAs designed for the knockout of *taok1* in zebrafish using the CRISPR-Cas9 system. A) *taok1*_sgRNA1 (target located in exon 4). B) *taok1*_sgRNA2 (target located in exon 6). Image generated by the author using Synthego's Verify Guide Design webtool v1.3²⁸².

The designed guides were ordered as single-strand DNA oligonucleotides, as indicated in Figure 126 (A-B). Those sequences were used to generate a 117 bp DNA-template of the sgRNA (Figure 126 C and Figure 127) that could be used for *in vitro* transcription to generate the full sgRNAs to be injected into zebrafish embryos.



Figure 126. Sequences of interest for the generation of sgRNAs by *in vitro* transcription. A) Single-strand DNA oligonucleotides with the target sequences (crRNA) selected for knocking out *taok1*. They also include the T7 promoter sequence and an annealing region complementary to the *tracrRNA* DNA-template. B) Single-strand DNA oligonucleotide that encodes the *tracrRNA* portion of a sgRNA for CRISPR-Cas9. C) Full length DNA templates (117 bp) of the designed sgRNAs for the knockout of *taok1*.

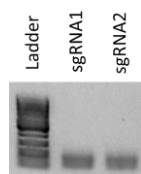


Figure 127. Synthesis verification of sgRNA DNA-templates (117 bp) in an agarose gel loaded with a 100 bp DNA ladder.

4.3.4.3. Characterization of the wild type individuals.

After guide selection, it was also necessary to design suitable primers to sequence the target regions of those guides. Since the designed guides do not target the same region, it was necessary to design two primer pairs, one for each. Those sequences, which were designed with Primer Blast, are indicated in Table 39.

Table 39. Sequencing primers for the identification of mutations in zebrafish in the target regions for the *taok1* knockout guides.

GENE	TARGET REGION	PRIMER SEQUENCE (5' - 3')	AMPLICON SIZE (bp)	ANNEALING (°C)
<i>taok1</i>	sgRNA 1 (Exon 4)	Forward: CATGGGTAAGAACTCAGCCCT Reverse: GGGAGTGGAGGTACGCCAAA	477	60 °C
<i>taok1</i>	sgRNA 2 (Exon 6)	Forward: GCCCATACTCTATCCTTGCGC Reverse: GGGAAGGATCGCCACTCCAC	688	60 °C

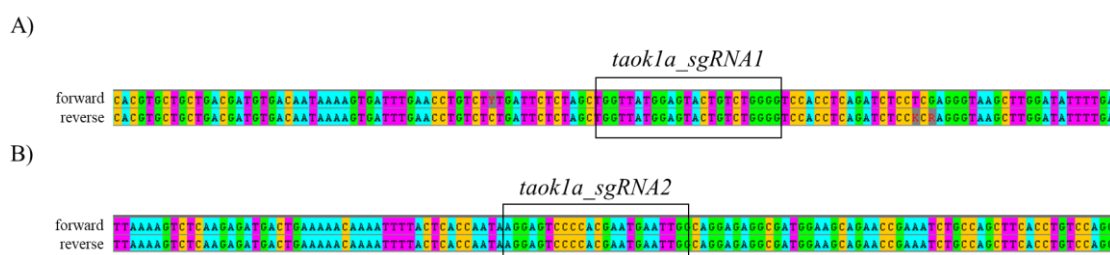


Figure 128. Wild type sequence of zebrafish in the target region for the selected knockout guides (in black boxes). A) Target region for *taok1*_sgRNA1. B) Target region for *taok1*_sgRNA2.

On the one hand, this was necessary to verify the WT sequence of the animals to avoid the presence of genomic variants that could interfere with the guide recognition of the target position, but also to detect the presence of genome editing events after CRISPR-Cas9. In Figure 128 can be observed that the sgRNA target regions do not show unexpected genomic variants.

4.3.4.4. CRISPR-Cas9 microinjection and guide efficiency analysis in the F0 generation of *taok1*-mutant zebrafish.

Once the sgRNA guide had been synthesized by *in vitro* transcription, they were co-injected with the Cas9 protein in zebrafish zygotes (F0 zebrafish generation). Two days later, several pools of injected embryos and WT controls were taken to be used in an heteroduplex analysis for the detection of CRISPR-mediated genome editing events.

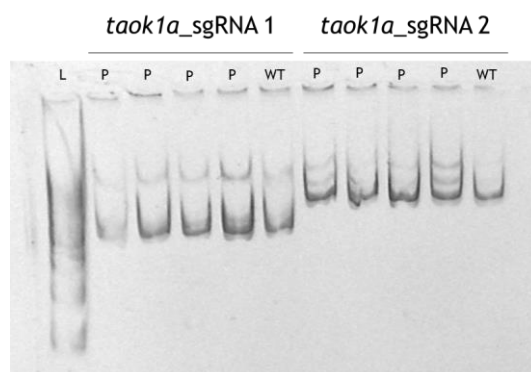


Figure 129. Heteroduplex analysis conducted in an acrylamide gel to verify guide efficiency. L: ladder 100 bp. P: pool of injected embryos from the F0. WT: pool of wild type non-injected embryos.

To do so, amplification products from the target regions of the selected sgRNAs were loaded in acrylamide gels and underwent electrophoresis. Briefly, this analysis is based on the detection of genetic variations or differences in DNA sequences (mutations, polymorphisms, etc.) due to their different mobility in an acrylamide gel. As can be seen in Figure 129, the migration of amplicons from the target region of *taok1*_sgRNA1 does not differ among the injected and WT embryos, suggesting that the guide might not be efficient. However, for the amplicons of the *taok1*_sgRNA2 target region it can be observed a different migration pattern between the WT and injected pools of embryos, which might indicate the presence of genome editing events in that region.

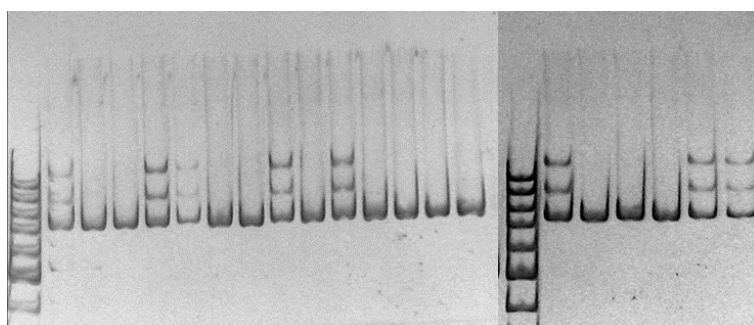


Figure 130. Heteroduplex analysis conducted in an acrylamide gel to genotype adult zebrafish from the F0 generation (n=20). Lanes 1 and 16 were loaded with 100 bp ladder, the remaining wells were loaded with amplicons from the target region of *taok1*_sgRNA2 obtained from different F0 individuals (ordered from 1 to 20, starting from the right).

Since at least one of the guides seemed to be efficient (*taok1*_sgRNA2), the remaining individuals from the F0 generation were maintained into adulthood, time by which they were genotyped. An heteroduplex analysis was conducted again using samples from the surviving adults from the F0 generation (n = 20) to detect genomic alterations in the target region for

taok1_sgRNA2 (Figure 130). The results of that analysis indicated that 8 individuals (1, 4, 5, 8, 10, 15, 19, and 20) were potentially edited, although they might be mosaics.

The 8 adults from the F0 generation identified as carriers of mutations in *taok1* were selected as the parental generation to obtain a homogeneous zebrafish line mutant for the gene of interest.

4.3.4.5. Obtention of zebrafish from the F1 generation and genetic characterization.

The first step to obtain a homogenous zebrafish mutant line is to isolate and characterize the different mutations in the gene of interest. To do so, the 8 adults from the F0 generation previously identified as harbouring mutations in *taok1* were selected and separated according to their sex, after which they were outcrossed with WT individuals (F0 males x WT females and F0 females x WT males). This strategy allows the isolation of mutated *taok1* alleles on different individuals from the offspring (F1 generation), with the objective of characterizing them and selecting those more likely to cause a *knockout* in the gene of interest. Therefore, the offspring from the F1 generation would be composed of WT and heterozygous zebrafish.

The F1 individuals were genotyped with a heteroduplex analysis, to identify the carriers of mutations in *taok1* (heterozygotes). As can be seen on Figure 131, only 7 individuals from the F1 reached adulthood, 4 of which were classified as heterozygotes: two males (1, 2) and two females (4 and 5).

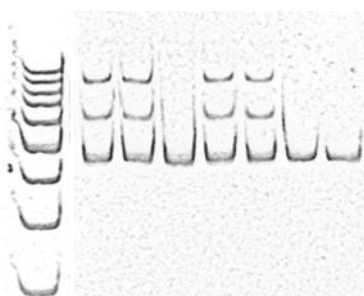


Figure 131. Heteroduplex analysis conducted in an acrylamide gel to genotype adult zebrafish from the F1 generation (n=7). A 100 bp ladder was loaded in lane 1, the remaining wells were loaded with amplicons from the target region of *taok1_sgRNA2* obtained from different F1 individuals (ordered from 1 to 7, starting from the right).

Next, the *taok1* alleles present in heterozygotes were isolated through bacterial cloning, after which they were sequenced to precisely identify mutations in the gene of interest. As can be observed in Figure 132, one allele carried an insertion of 4 nucleotides (5'-GTTA- 3'). This frameshift mutation originates a new stop codon in the sequence. In consequence, a peptide of 230 aa is generated, disturbing the protein sequence inside the kinase domain.



Figure 132. Alleles identified through bacterial cloning in the F1 generation. Sequences were aligned with the reference sequence of *taok1* exon 6 (indicated as Refseq *taok1*), the target region of *taok1_sgRNA2*. One of the sequences carries an insertion of 4 bp (in the middle). The new stop codon is marked in black.

4.3.4.6. Obtention of zebrafish from the F2 generation and phenotypic characterization.

The adults from the F1 generation were crossed twice. The first time, F1 adults were crossed among themselves to produce a F2 generation, composed of knockout individuals (25 % carrying two altered copies of *taok1*), heterozygotes (50 % carrying one altered copy of *taok1*) and WT (25 % without mutations in *taok1*). Unfortunately, on the period between crosses, both heterozygous females from the F1 generation died, only leaving surviving males in the F1. For this reason, F1 males were crossed with WT females. Therefore, the expected offspring would be composed of 50 % WT animals and 50 % heterozygotes, but zebrafish carrying two altered copies of *taok1* would not be produced.

The offspring obtained from those crosses was genotyped using fragment analysis, to rapidly identify the alleles present of each sample based on their size. An example of the identification of the different genotypes using this technique can be found in Figure 133. The number of individuals obtained in the F1 crosses, classified by their genotype, is shown in Table 40.

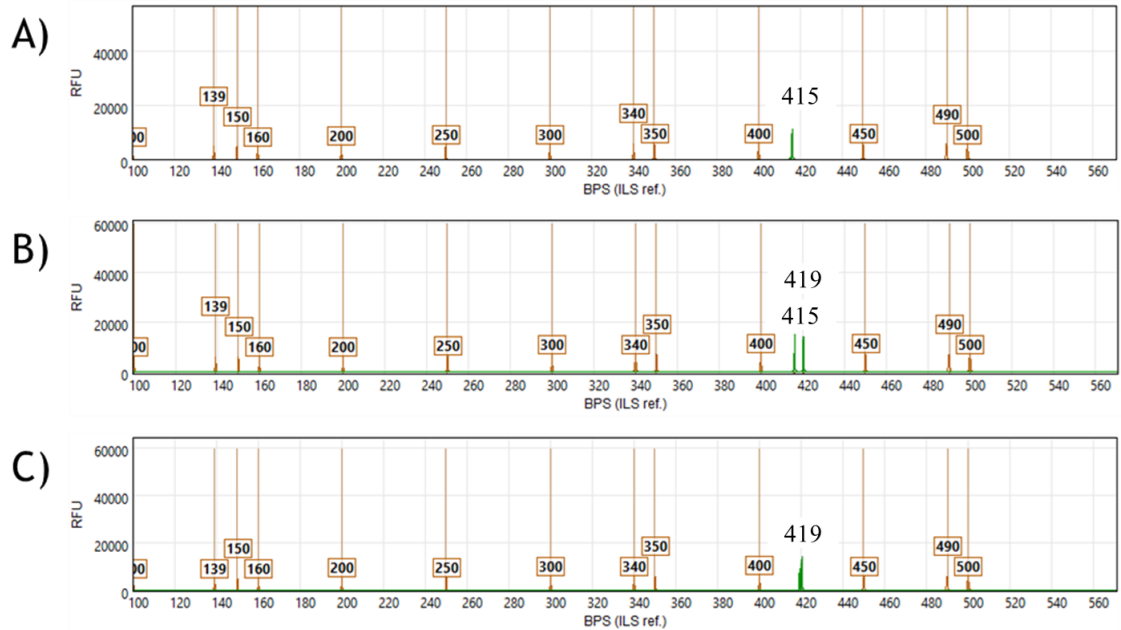


Figure 133. Representation of the types of electropherograms obtained after fragment analysis for the genetic characterization of zebrafish individuals (ladder in orange, samples in green). A) Homozygous individual with two wild type alleles (415 bp). B) Heterozygous individual with one wild type (415 bp) and one edited (419 bp) allele. C) Homozygous individual with two edited alleles (419 bp).

Table 40. Genotypes of the offsprings obtained after crossing F1 individuals.

GENOTYPE	OFFSPRING 1 st CROSS	OFFSPRING 2 nd CROSS
Wild type (<i>taok1</i> ^{415/415})	8	19
Heterozygote (<i>taok1</i> ^{415/419})	22	20
Knockout (<i>taok1</i> ^{419/419})	7	-

4.3.4.6.1. Morphometric analysis of *taok1* mutant zebrafish.

Morphometric analysis of the F2 individuals was conducted at 48 and 96 hpf, by measuring parameters of interest such as the body length, the head size, and the eye size. By doing so, it was intended to identify if the normal zebrafish development was altered due to

mutations in *taok1*, both in homozygosis and heterozygosis. Figure 134 shows representative images of individuals from the F2 generation at 48 hpf and 96 hpf.

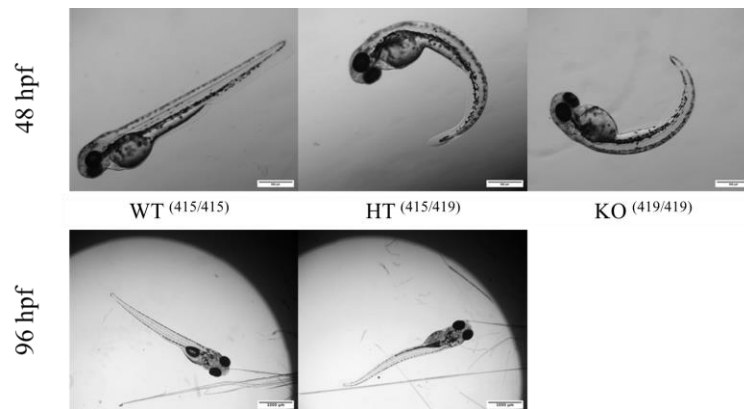
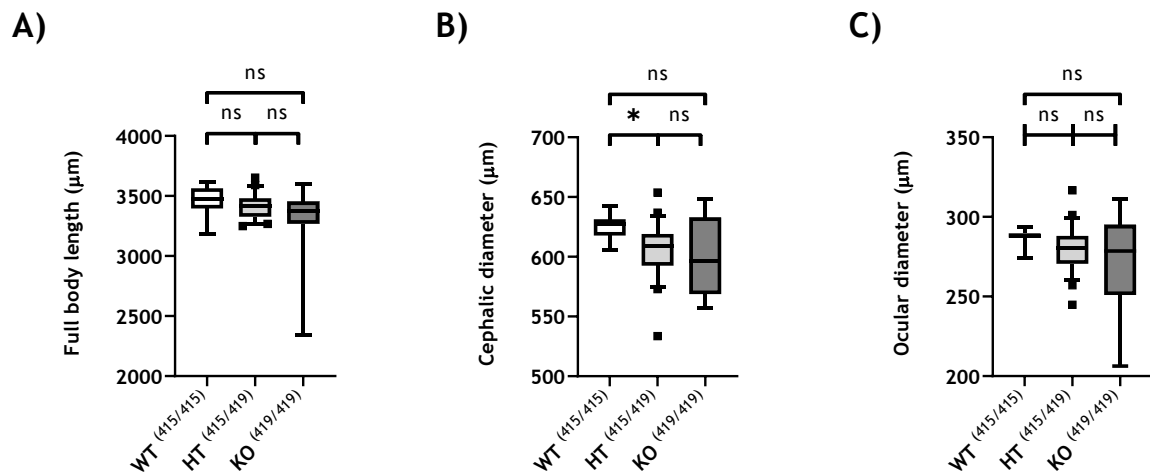


Figure 134. Representative images of individuals from the F2 generation at 48 hpf and 96 hpf, classified by their genotype.

Table 41. Morphometric measurements of the F2 generation at 48 hpf.

PARAMETER	EXPERIMENTAL CONDITION	MEAN	SD
Body length (μm)	Wild type (WT ^{415/415})	3455	134.2
	Heterozygous (HT ^{415/419})	3415	110.4
	Knockout (KO ^{419/419})	3245	411.8
Head diameter (μm)	Wild type (WT ^{415/415})	625.4	11.08
	Heterozygous (HT ^{415/419})	605.5	25.04
	Knockout (KO ^{419/419})	599.3	34.26
Eye diameter (μm)	Wild type (WT ^{419/415})	287.3	5.682
	Heterozygous (HT ^{415/419})	279.8	15.04
	Knockout (KO ^{419/419})	271.8	34.57



COMPARED GROUPS	FULL BODY LENGTH	CEPHALIC DIAMETER	OCULAR DIAMETER
WT ^{415/415} vs HT ^{415/419}	0.730	0.015	0.160
WT ^{415/415} vs KO ^{419/419}	0.443	0.199	0.487
HT ^{415/419} vs KO ^{419/419}	0.560	0.898	0.573

Figure 135. Morphometric analysis of the F2 generation at 48 hpf. Analysis conducted with wild type individuals (WT ^{415/415}), heterozygous individuals that carry the knockout mutation in one allele (HT ^{415/419}) and homozygous individuals that carry the mutations in both alleles (KO ^{419/419}). A) Full body length. B) Cephalic diameter. C) Ocular diameter. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

The mean and standard deviation of the morphometric parameters addressed in this section at 48 hpf are recollected in Table 41, while the results of the statistical analysis can be found in

Figure 135. Apart from the smaller cephalic size of the heterozygous (HT ^{415/419}) individuals when compared to the wild type (WT ^{415/415}), no significant differences were found in body length, head size or eye size in the individuals from the F2 generation when categorized by the presence of mutations in *taok1*. Nevertheless, it is worth mentioning that a slight tendency can be observed in the data, which suggests a decrease in body length, head size and eye size associated to the presence of mutant *taok1* alleles.

To complete the morphometric analysis at 48 hpf two ratios of interest were quantified: the ratio between the head and body, and the ratio between the eye and head of the embryos (Table 42). As can be seen in Figure 136, no significant differences were found in none of those ratios in the F2 embryos at 48 hpf.

Table 42. Morphometric ratios of the individuals from the F2 generation at 48 hpf.

PARAMETER	EXPERIMENTAL CONDITION	MEAN	SD
Ratio: head/body [Cephalic diameter (μm)/full body length (μm)]	Wild type (WT ^{415/415})	0.181	0.009
	Heterozygous (HT ^{415/419})	0.178	0.009
	Knockout (KO ^{419/419})	0.187	0.024
Ratio: eye/head [Cephalic diameter (μm)/full body length (μm)]	Wild type (WT ^{415/415})	0.460	0.012
	Heterozygous (HT ^{415/419})	0.463	0.027
	Knockout (KO ^{419/419})	0.453	0.046

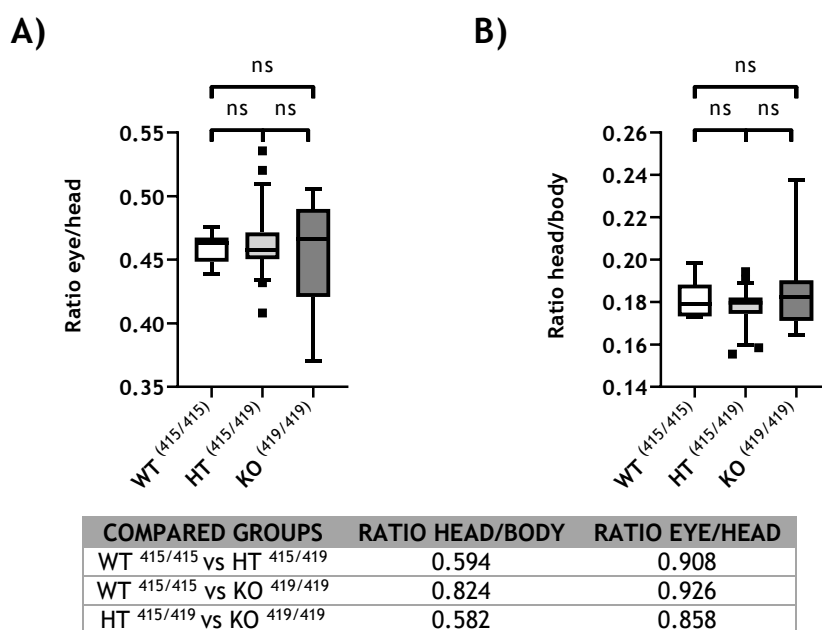


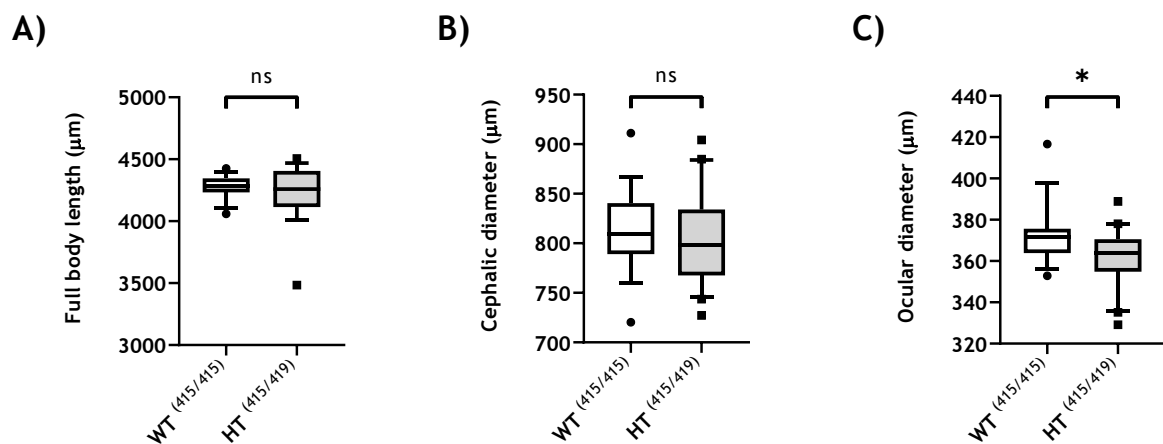
Figure 136. Morphometric analysis of the F2 generation at 48 hpf. Analysis conducted with wild type individuals (WT ^{415/415}), heterozygous individuals that carry the knockout mutation in one allele (HT ^{415/419}) and homozygous individuals that carry the mutations in both alleles (KO ^{419/419}). A) Ratio between the ocular diameter and the cephalic diameter of the individuals. B) Ratio between the cephalic diameter and the full body length of the individuals. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

As mentioned earlier in this section, morphometric parameters were also studied at 96 hpf. The mean and standard deviation of the data can be observed in Table 43, while the statistical analysis is represented in Figure 137. Smaller eyes were identified at 96 hpf in heterozygous

(HT^{415/419}) individuals, but no major differences were found in body length or head size when compared to the wild type (WT^{415/415}), although the tendency of the data might suggest a slight decrease in both parameters. Like it happened at 48 hpf, no significant differences were found in the head/body ratio, nor in the eye/head ratio (Table 44, Figure 138).

Table 43. Morphometric measurements of the F2 generation at 96 hpf.

PARAMETER	EXPERIMENTAL CONDITION	MEAN	SD
Body length	Wild type (WT ^{415/415})	4275	97.81
	Heterozygous (HT ^{415/419})	4228	230.5
Head diameter	Wild type (WT ^{415/415})	813.4	43.37
	Heterozygous (HT ^{415/419})	806.3	47.01
Eye diameter	Wild type (WT ^{415/415})	373	14.44
	Heterozygous (HT ^{415/419})	361.5	14.27



COMPARED GROUPS	FULL BODY LENGTH	CEPHALIC DIAMETER	OCULAR DIAMETER
WT ^{415/415} vs HT ^{415/419}	0.419	0.629	0.017

Figure 137. Morphometric analysis of the F2 generation at 96 hpf. Analysis conducted with wild type individuals (WT^{415/415}) and their heterozygous siblings that carry the knockout mutation (HT^{415/419}). A) Full body length. B) Cephalic diameter. C) Ocular diameter. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

Table 44. Morphometric ratios of the individuals from the F2 generation at 96 hpf.

PARAMETER	EXPERIMENTAL CONDITION	MEAN	SD
Ratio: head/body	Wild type (WT ^{415/415})	0.191	0.008
[Cephalic diameter (μm)/full body length (μm)]	Heterozygous (HT ^{415/419})	0.191	0.014
Ratio: eye/head	Wild type (WT ^{415/415})	0.459	0.021
[Cephalic diameter (μm)/full body length (μm)]	Heterozygous (HT ^{415/419})	0.449	0.021

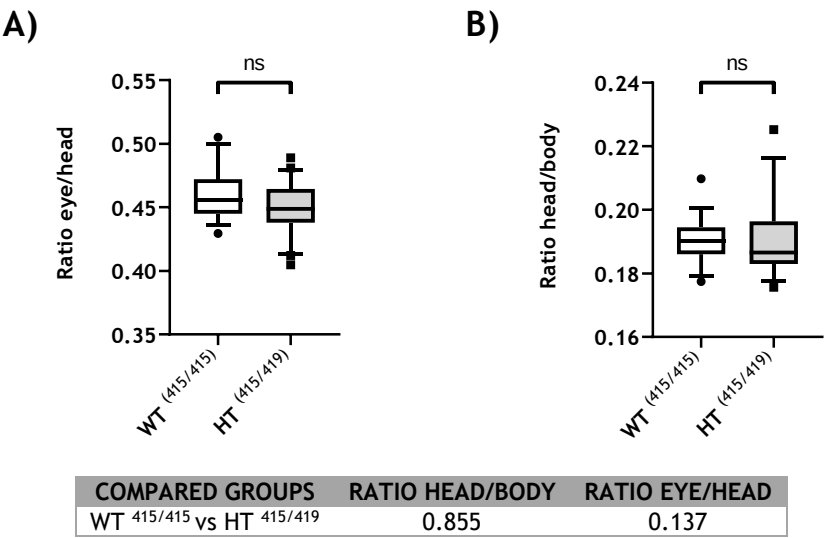


Figure 138. Morphometric analysis of the F2 generation at 96 hfp. Analysis conducted with wild type individuals (WT ^{415/415}) and heterozygous individuals that carry the knockout mutation in one allele (HT ^{415/419}). **A)** Ratio between the ocular diameter and the cephalic diameter of the individuals. **B)** Ratio between the cephalic diameter and the full body length of the individuals. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

In summary, morphometric analyses at 48 hpf and 96 hpf suggest a slight reduction in body, head and eye sizes associated with alterations in *taok1*, but those differences are not big enough to be of statistical significance. Taken together, these results might indicate that alterations in *taok1* are not associated with a strong developmental delay in zebrafish.

4.3.4.6.2. Behavioural response of the F2 larvae using the light-dark test.

The light-dark test was used to study the presence of alterations in the anxiety-like to illumination changes between individuals from the F2 generation.

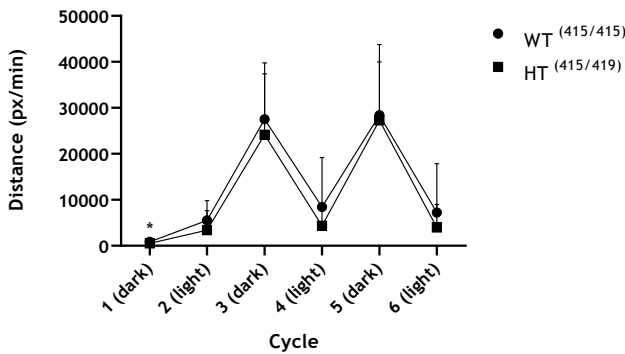


Figure 139. Light-dark response of 96 hfp individuals from the F2 generation. Analysis conducted with wild type individuals (WT ^{415/415}) and heterozygous individuals that carry the knockout mutation in one allele (HT ^{415/419}). Movements were measured as distance (in pixels) travelled per minute. Data represented as mean + SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

The distance travelled by each zebrafish individual was measured in a tracking system that automatically altered light and dark cycles. The result of this test revealed that the anxiety-like response is overall similar in WT and heterozygous zebrafish for the *taok1* knockout mutation of interest. However, heterozygous individuals seem to be hypoactive in light cycles in comparison to the WT controls (Figure 139, Figure 140).

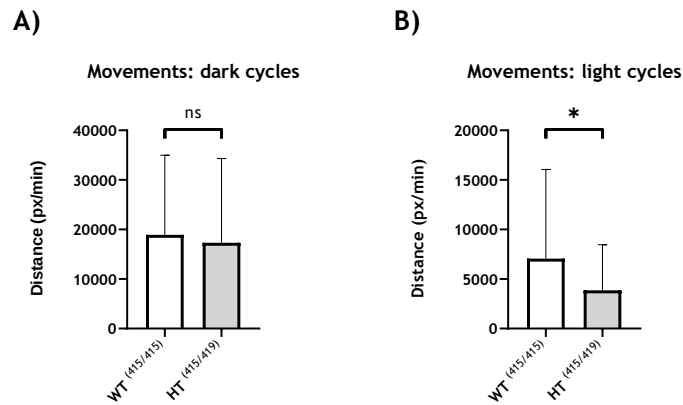


Figure 140. General representation of the movements of the individuals from the F2 generation (96 hfp) in light and dark cycles. Analysis conducted with wild type individuals (WT ^{415/415}) and heterozygous individuals that carry the knockout mutation in one allele (HT ^{415/419}). A) Movements detected in light cycles. B) Movements detected in dark cycles. Data represented as mean + SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.

DISCUSSION

5. DISCUSSION

The results of this thesis will be discussed in the same order they were presented. First, the results derived from *in vitro* models of *BTBD8* will be discussed, paying especial attention to interpret the obtained proteomic data in the context of neurodevelopmental alterations. Second, the observations conducted in the *fosl2* knockdown model will be compared to the available information about *FOSL2* role in developmental and the suitability of using zebrafish as a model for its study will be addressed. Third, the findings made in *TAOK1* *in vitro* and *in vivo* models will be commented, discussing the scRNA-seq preliminary data and zebrafish phenotype in the context of ASD. A final section was included to discuss the suitability and limitations of the selected experimental approaches, since the same strategy was followed to generate hiPSCs based models to study *BTBD8* and *TAOK1* *in vitro*.

5.1. GENERATION AND CHARACTERIZATION OF *IN VITRO* MODELS TO STUDY THE FUNCTION OF *BTBD8*.

As a part of the present doctoral thesis, a human *in vitro* model based on hiPSCs was generated to study *BTBD8*'s role in neurodevelopment. *BTBD8*, a member of the BTB/POZ domain-containing protein family, is thought to be involved in critical cellular functions, including transcriptional regulation and synaptic transmission.

Understanding the intricate mechanisms governed by *BTBD8* is essential for unravelling the complexities of normal cellular processes and how their dysregulation can lead to diseases and disorders development. However, scarce publications have addressed *BTBD8*'s role in any of those process. More specifically, just one publication has conducted functional studies to validate the predicted function of the gene at the moment. In that publication, Piccini *et al.*²¹¹ silenced *BTBD8* (previously annotated as *KIAA1107*) with short hairpin RNAs (shRNAs) on murine primary cultures. Their work highlighted the protein's role in synaptic vesicle recycling, and thus synaptic transmission although no more studies have been conducted to neither validate nor extend their findings.

The model generated in this thesis is, as far as we know, the first human *in vitro* model generated to study *BTBD8*. Moreover, the *BTBD8*-mutant model was generated in hiPSCs making it suitable for studying its role in multiple contexts, including cell fate determination, tissue-specific functions, and potential implications in disease states affecting diverse cell-types. Models based on hiPSCs, therefore, are a powerful tool for unravelling the complexity of *BTBD8* biology and hold immense promise for advancing both basic research and clinical applications.

In essence, studying *BTBD8* is not only crucial for expanding our knowledge of basic biology and disease pathology, but also for the potential development of novel therapeutic strategies aimed at addressing conditions associated with its dysfunction. In this context, the *BTBD8* *in vitro* model developed as a part of this thesis has a great potential to uncover the suitability of *BTBD8* as a potential diagnostic or prognostic marker, as well as to test novel therapeutical approaches to ameliorate *BTBD8*-associated molecular implications.

In the following sections, the *BTBD8*-associated findings reported in this thesis will be interpreted and discussed.

5.1.1. Functional implications of BTBD8's expression pattern and localization.

Piccini *et al.* studied the expression of *BTBD8* in mouse brain, where they found the highest levels in specific cerebral regions: cortex, hippocampus, and striatum. In addition, they reported *BTBD8* levels increased over time from embryonic day 18 until postnatal day 30, after which they decreased²¹¹. In concordance to this, data from this thesis points out to a low *BTBD8*'s expression in hiPSCs-derive brain organoids, including samples from day 45 to 90 (Figure 72). Additionally, as observed in 2D cultures, *BTBD8* detection is reduced as the days of differentiation increased (Figure 65).

According to public databases *BTBD8*'s subcellular localization includes both the nucleus and the cytoplasm. Xu *et al.* have hypothesized that the protein was found in the nucleus of foetal-origin cells (such as HEK293) while in adult cells, like NIH/3T3, it was located in the cytoplasm²¹⁰. Nevertheless, it is important to notice that those cell lines are derived from different species, human and rat, respectively. Also, *BTBD8* is localized in the nucleus of the human cancerous cell lines such as A-431 (epidermoid carcinoma), U-251MG (glioblastoma), and U2OS (osteosarcoma), as indicated by The Human Protein Atlas database²⁰⁷. In addition, Piccini *et al.* have reported that *BTBD8* localizes in the axonal processes and synaptic terminals of primary cortical cultures from mice. Finally, Lipovich *et al.* showed another interesting insight into *BTBD8* expression, as it was one of the 13 *loci* differentially expressed between *Homo sapiens* and *Gorilla gorilla*. As indicated by the authors, *BTBD8* is mainly expressed in the foetal brain in humans, while in gorillas *BTBD8* mRNA could be recovered from adult brain. Lipovich *et al.* also suggest that the observed differences could be explained by a loss of a splicing event that only affected *Homo sapiens*³⁰².

Results obtained in this thesis indicate that *BTBD8* is detected mainly in the nucleus of hiPSCs-derived neuronal cells (Figure 65 and Figure 72). According to the hypothesis stated by Xu *et al.*²¹⁰, these results could be explained by the general immaturity of the neuronal cultures, being *BTBD8* located in the nucleus of precursor cell. However, in human brain organoids the subcellular localization of *BTBD8* does not seem to change from day 45 to day 90. Another possible explanation to these observations would be that *BTBD8* subcellular localization differs between species. Limited information about *BTBD8* localization and function make it imperative to perform further studies to draw conclusive insights into this issue, such as conducting additional immunostainings in other cell lines from different species.

5.1.2. Morphological alterations in Ctrl.BS-PBiPS37-SvF4-1-*BTBD8*-i2 brain organoids.

The developmental tracing of human brain organoids derived from the hiPSCs line Ctrl.BS-PBiPS37-SvF4-1-*BTBD8*-i2 revealed a significant increase in size in comparison to isogenic controls (Figure 62). While clinical reports detailing *BTBD8*-associated phenotypes in humans are currently absent, this finding strongly suggests a potential macrocephaly phenotype associated with *BTBD8* mutations.

Macrocephaly is a feature commonly identified in individuals diagnosed with ASD^{35,303–305}. In fact, a meta-analysis conducted by Sacco *et al.* revealed that 15.7 % of ASD cases shared this morphological trait, contrasting with only 3% of controls³⁰⁶. Although not universally present, the occurrence of macrocephaly suggests a connection between brain structure and ASD. In this sense, several publications in the field have associated macrocephaly with an altered pattern of cerebral growth, which could be potentially caused by an imbalance between proliferation and differentiation of neural progenitors^{35,307–309}. Supporting this hypothesis,

mutations in genes implicated in cell-cycle control, such as *ANKRD11*³¹⁰ and *PTEN*³¹¹, had been found in ASD patients.

Unfortunately, phenotypical data is not fully specified in the patient that carries out mutation of interest, and no information is available concerning the presence of macrocephaly. Hence, it is imperative to improve multidisciplinary among clinical care providers, genetic experts, and translational biomedical researchers in order to advance towards real personalized medicine.

Therefore, the molecular mechanisms through which *BTBD8* dysfunction could lead to macrocephaly were not explored in this thesis, but open interesting lines of investigation for the future, even more if combined with thorough phenotypic and clinical information of the affected individuals.

5.1.3. Interpretation of proteomic data from *BTBD8*-mutant neuronal cultures in the context of ASD.

The impact of endosomal alterations in ASD's development has been known for years, as reviewed by Patak *et al.*³¹². The dysregulation of the endosomal system can lead to a cascade of events influencing gene expression, synaptic transmission, cell signalling, and cytoskeleton remodelling, among others. However, the intricate interplay between genes regulated by the endosomal system and those governing its function has not been fully uncovered. In this context, clarifying the function of *BTBD8* could provide valuable insights into the intricate mechanisms contributing to a correct neural function, but also how its alterations can lead to ASD.

In 2017, Piccini *et al.* identified *BTBD8* as a component of the clathrin-mediated endocytosis pathway²¹¹. Despite being able to assemble on their own, clathrin-coated vesicles need adaptor proteins (AP) to interact with the plasma membrane. *BTBD8* interacts with one of those adaptor protein complexes, known as AP-2. The proteomic analysis conducted in this thesis identified two subunits of the AP-2 complex (AP2A1 and AP2B1) depleted in *BTBD8*-mutant cultures differentiated for 40 days. AP2A1 encodes the AP-2 subunit α , responsible of membrane binding, while AP2B1 encodes a β subunit responsible of clathrin binding³¹³. The depletion of both subunits correlates with *BTBD8*'s reduction, which is consistent with reports of AP-2 degradation after *BTBD8*'s knockdown²¹¹.

Piccini *et al.* also predicted *BTBD8* to interact with another AP complex, AP-3, although they did not experimentally test the interaction²¹¹. While AP-2 expression is required for neural maturation, the neuron specific AP-3 is thought to mediate immature synaptic vesicle recycling, along with inter-organelle transport^{314,315}. As mentioned in the results, AP-3 was only detected in *BTBD8*-mutant cultures differentiated for 60 days (AP3B2 subunit represented in Figure 79C). This observation resembles the findings from Shetty *et al.*³¹⁶, who generated a deficient mouse model for the AP-2 interactor NCAM that also showed accumulation of AP-3 complexes. In NCAM deficient neurons, AP-3 accumulation was accompanied by impaired maturation of the synaptic endocytic machinery, which might be also happening in *BTBD8*-mutant cultures. The absence of AP-3 in *BTBD8*-mutant cultures differentiated from 40 days could be related to the enrichment in NSC-markers observed in Figure 78.

Another AP-2 interactor was found altered in *BTBD8*-mutant cultures: SNAP91 (Figure 79D)³¹⁷. SNAP91, which participates in the linkage between clathrin and receptors, was depleted in day 40 *BTBD8*-mutant cultures. This finding further supports the hypothesis of *BTBD8* alterations leading to impairments in clathrin-mediated endocytosis. In addition, the synaptic phenotypes observed after *BTBD8* knockdown²¹¹ and SNAP91 depletion³¹⁸ are very

similar (enlarged synaptic vesicles in low densities). It is very probable that neurons derived from the mutant model developed in this thesis also shared this phenotype, although it remains to be analysed.

Also, as reviewed by Robinson *et al.*, AP complexes binding to the plasma membrane is dependent on phosphoinositides (PI)³¹⁹, which biosynthesis seem to be altered in the *BTBD8*-mutant model according to the proteomic data. Phosphatidylinositol 4,5-bisphosphate (PIP₂) and phosphatidylinositol 3,4,5-trisphosphate (PIP₃) are important regulators of cellular functions, including growth, metabolism, and transport. Tariq *et al.* have an extensive review on which they addressed the impact of alterations in PI biosynthesis on neurodevelopmental and neurodegenerative disorders, including ASD³²⁰.

Proteomic data also revealed alterations in DNM2 (Figure 79J), exclusively detected in *BTBD8*-mutant cultures, where its expression increased over time. DNM2 is a GTPase that binds and hydrolyses GTP as a force-producing mechanism necessary for cytoskeleton rearrangements (microfilaments and microtubules). DNM2 function not only has been associated to the regulation of neuronal morphology (axon growth), but also to clathrin dependent and independent endocytosis^{321,322}. Despite the absence of reported direct interaction, the functional similarities between DNM2 and BTBD8 suggest their involvement in common pathways within the CNS. *DNM2*-related phenotypes are also very similar to the ones observed in the *BTBD8*-mutant model generated in this thesis. Heterozygous mutations in *DNM2* have been linked to Charcot-Marie-Tooth type 2, an axonal neuropathy that affects the peripheral nervous system. Also, heterozygous mutations in *DNM2* are associated to autosomal dominant myopathies characterized by alterations in the ubiquitin-proteasome pathway, autophagy, and DNM2 accumulation inside the cells, all of which were also found in the *BTBD8*-mutant cultures^{321,323–325}. Interestingly, both increases and decreases in autophagy have previously reported in ASD cases. As reviewed by Dana *et al.*³²⁶, autophagy is thought to be key for a proper NSCs differentiation, excitatory/inhibitory neurotransmission, dendrite and axonal growth, and synapse formation, all of which seem to be altered in the generated *BTBD8*-mutant model. Moreover, *DNM2* knockout has been linked to lethal phenotypes in diverse species, including *Mus musculus* and *Homo sapiens*^{321,327}. This prompts the question: is a similar lethality associated with *BTBD8* mutations? Examining the CRISPR-Cas9 experimental outcomes, it is remarkable that none of the 44 isolated colonies contained stop gain mutations in both *BTBD8* alleles, despite the high efficiency of the selected CRISPR-Cas9 guide. The selected *BTBD8*-mutant line (Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8-i2*) carries two mutant alleles, one containing a stop gain mutation and the other 3 nts deletion. This combination could potentially result in either a *BTBD8* full knockout or haploinsufficiency, depending on the severity of the 3D alterations associated with the allele lacking one amino acid. Contrary to the expectation of a full knockout, BTBD8 was detected in organoids derived from the mutant hiPSC-line, suggesting that alterations in BTBD8 protein do not induce its degradation within the cells. Considering all this information, it is more likely that Ctrl.BS-PBiPs37-Sv4F-1-*BTBD8-i2* exhibits haploinsufficiency, sufficient to cause phenotypical alterations in the cultures. This aligns with Alonso *et al.*'s findings of heterozygous mutations in *BTBD8* associated with ASD^{26,216}. To date, there are no reports of the phenotypic consequences of a full knockout of *BTBD8* in humans, raising the possibility of lethality. In contrast, Piccini *et al.* observed no viability alterations when they knockdown *BTBD8* in mouse primary cortical neurons, although morphological alterations were reported²¹¹. Nevertheless, this observation could be associated to the maturity of the cells. As it has been previously mentioned, *BTBD8* expression and domain composition suggest an involvement in early development. Thus, a full knockout of *BTBD8* could be lethal both in embryos and stem cells, but do not affect the

viability of mature cell lines. To test this hypothesis, a comprehensive RNA-seq analysis is now being developed to compare the transcriptomic profiles of wild-type and *BTBD8*-mutant hiPSCs.

DNM2 was not the only dynamin altered in the proteomic data. DNM1, a brain specific GTPase was also altered (Figure 79I). While wild type cortical neurons display rather stable levels of DNM1, the protein appeared to be slightly dysregulated in mutant cells. As indicated by Yamashita *et al.*, DNM1 localizes in clathrin-coated pits in the plasma membrane, where it contributes to the energy requirements needed for membrane fission³²⁸. Thus, alterations in *BTBD8*-mutant cultures could be associated with impairments in clathrin-mediated endocytosis. Considering its function, alterations in DNM1 could also be linked to the alterations in purine catabolism detected in the proteomic analysis.

The *MAPT* derived protein, tau, plays a crucial role in establishing neuronal polarity and is essential for the interactions between microtubules and the plasma membrane. Results presented in this thesis demonstrate a progressive increase in *MAPT* expression in WT neuronal cultures (Figure 80H). The same tendency is observed in *BTBD8*-mutant cultures, although tau is reduced in day 40 and increased in day 60 when compared to the control, which was also observed in *BTBD8*. In relation to this, Jiang *et al.* found alterations in *BTBD8*'s expression in response to *MAPT* mutations, indicating a potential interaction between both, although they did not further explore this hypothesis²⁹⁹. These authors reported a decrease in *BTBD8* expression in hiPSCs-derived neurons and human brain samples carrying MAPT p.R406W, a mutation link to frontotemporal lobar degeneration with tau inclusions (FTLD-tau). In addition, decrease in GABA receptors was observed in these models. Interestingly, a reduction in the GABAergic population was also observed in the *BTBD8*-mutant organoid model generated in this thesis (Figure 67, Figure 68). The protein tau is mainly known for its role in neurodegenerative diseases, where its phosphorylation status has been linked to dysregulation of GABA receptors^{329,330}. In addition, alterations in GABA neurotransmission early on development can have consequences in neural progenitor cell differentiation, elongation of neurites and synapse formation, thus leading to neurodevelopmental disorders²⁹⁹. However, the interplay between tau and GABA in neurodevelopment is poorly understood. Some studies, such as those by Chang *et al.*, suggest tau involvement in disorders characterized by neuronal network alterations, including ASD³³¹. Reduced tau levels have been identified in ASD models with impaired excitatory/inhibitory balance and altered synchrony of the neuronal networks. In addition, Tai *et al.* reported an improvement in the phenotype of ASD mouse models with epilepsy (*Scn1a*^{RX/+} and *Cntnap2*^{-/-}) upon tau removal. Furthermore, tau reduction is associated with potentiated PI3K/Akt signalling, which relevance in ASD development has been addressed earlier on this discussion³³².

STX8, a syntaxin that interacts with tau interactor and participates in vesicle trafficking³³³, is also altered in the proteomic analysis. STX8 is absent from *BTBD8*-mutant cultures, further supporting the association between *BTBD8* and tau, which could be mediated by STX8 (Figure 79G, Figure 80H).

Another syntaxin is also found altered in *BTBD8*-mutant cultures: STX7 (Figure 79H). This syntaxin mediates neurotransmitter release and synaptic vesicle recycling, so its depletion in mutant cultures is another indication of endocytic impairments associated with *BTBD8* mutations. In epilepsy, reduced levels of STX7 have been found associated with GABA reductions, decreased excitatory/inhibitory ratios, and neurotransmitter release impairments^{334,335}. The impact of *BTBD8* knockdown in synaptic transmission has been reported by Piccini *et al.*²¹¹, but it remains to be addressed in the *BTBD8*-mutant model developed in this thesis.

STXBP1 levels were also studied in the proteomic data, since it participates in synaptic vesicle docking and fusion by interacting with GTPases and syntaxins³³⁶. Results from this thesis show reduced levels of STXBP1 in day 40 *BTBD8*-mutant cultures (Figure 79E), which resemble alterations found in AP-2 subunits (Figure 79A, B) and suggest impaired clathrin mediated endocytosis in *BTBD8*-mutant neurons. *De novo* variants in *STXBP1* have been associated with early onset of neurocognitive conditions such as early infantile epileptic encephalopathy type 4 (featuring developmental delay and ASD) and Rett syndrome, which no longer is classified as a subtype of ASD, but shares some of its characteristics^{337,338}. Interestingly, Miyamoto et al found that pharmacological normalization of synaptic transmission with ampakine had positive effects ameliorating the aggressive phenotype in *STXBP1*^{-/+} mice³³⁹. These findings support the potential of pharmacological targeting of synaptic vesicle-associated proteins to ameliorate ASD-symptoms. It is also worth noting that *STXBP1* knockout lead to perinatal neurodegeneration and, in mice, early mortality, further supporting the hypothesis of a lethal phenotype associated with *BTBD8* knockout^{336,337}.

Jiang *et al.* also reported colocalization of BTBD8 with SYT1²⁹⁹, a protein that regulates interactions between synaptic vesicle and the plasma membrane. SYT1 is also proposed to act as a calcium sensor that mediates neurotransmitter release. Interestingly, SYT1 is only detected in day 60 *BTBD8*-mutant proteomic data (Figure 79F). According to Wang *et al.*, overexpression of SYT1 could cause alterations in the time of membrane fusion and neurotransmitter release³⁴⁰. Tau *et al.* have also recently reported that SYT1 could be potentially associated with AMPA receptor exocytosis, thus altering the excitatory/inhibitory balance in the brain³⁴¹.

Piccini *et al.* also predicted DCTN1 (Figure 80A) as a BTBD8-interactor²¹¹. DCTN1 participates in axonal transport and synaptic vesicle recycling and its expression increases over time along with neurite formation, as reported by Boumil *et al.*^{342,343}. Reduced DCNT1 levels observed in day 40 *BTBD8*-mutant cultures could be associated with impairments in dendrite formation and synaptogenesis.

Apart from analysing the expression of genes related to endosomal pathways, the proteomic analysis presented in this thesis also include genes associated with protein degradation, transcriptional regulation, and cell signalling, which association with *BTBD8* and ASD will be discussed next.

BCLAF1, another potential interactor of BTBD8 as predicted by Piccini *et al.*²¹¹, acts a death promoting transcriptional repressor which is mediated by its interaction with BCL2. Proteomic data presented in Figure 80D indicate a strong presence of BCLAF1 in day 40 wild type cultures, which is reduced in day 60. On the contrary, BCLAF1 is absent from day 40 *BTBD8*-mutant cultures and strongly expressed at day 60. As indicated by Horiuchi *et al.*³⁴⁴, BCLAF1 depletion is associated with reduced levels of splicing and cell proliferation by mediating cell-cycle arrest in G2, while its overexpression is associated with induced apoptosis. Apoptosis plays an important role in brain development, and its dysregulation is oftentimes associated with neurodevelopmental disorders, including ASD, as reviewed by Wei *et al.*³⁴⁵. Interestingly, post-mortem brain samples of ASD patients usually show BCL2 reductions (30--50 % depending on the brain region), which are thought to be mediated by impairments in the PI3K/Akt pathway, also altered in *BTBD8*-mutant neurons as discussed earlier^{346,347}. In addition, oxidative stress and glutamate excitotoxicity also mediate apoptosis during brain development^{345,348}. Proteomic data obtained in this thesis show impairments in glutamate clearance which could potentially affect neural apoptosis mediated by BCL2.

Another potential interactor of BTBD8 is AHNAK²¹¹, a scaffold protein that participates in neural differentiation, migration, calcium homeostasis, and serotonergic

neurotransmission^{349,350}. Lu *et al.* reported that NSCs differentiation can be impaired by the downregulation of FAM107A, a protein that acts downstream AHNK³⁵¹. The increased levels of progenitor markers observed in *BTBD8*-mutant cultures (Figure 78) could be associated with AHNK absence (Figure 80G) and thus, impairments in NSCs differentiation. In addition, Sundararaj *et al.* experiments indicate that *AHNK* downregulation could be involved in depression behaviour, a common comorbidity to ASD.

Piccini *et al.*, also identified ATP6V1A, a vascular ATPase component, as a potential *BTBD8* interactor²¹¹. Guerrini *et al.* were able to associate alterations in this ATP-dependent pump with a developmental and epileptic encephalopathy phenotype, characterized by pH alterations in the lysosome leading to an increased accumulation of substrates and the development of brain abnormalities³⁵². As mentioned in the results section, the levels ATP6V1A were slightly reduced in day 60 *BTBD8*-mutants in comparison to the wild type control (Figure 80B), which can contribute to impairments in lysosome functioning and altered neuronal morphology.

Potential *BTBD8* interactors were also identified using STRING v12.0 webtool²⁹⁰. The only interactor identified in the proteomic data using this approach was CUL3. CUL3 is part of the cullin-RING-based BCR E3 ubiquitin ligase complex, and thus participates in the proteasomal degradation pathway. As shown in Figure 80C, CUL3 levels are increased in *BTBD8*-mutants at both time points in comparison to the wild type control. CUL3 alterations have been reported in a neurodevelopmental disorder with/without ASD and seizures. Interestingly, CUL3 interacts with KBTBD8, a protein that contains one BTB/POZ domain — *BTBD8* contains two as represented in Figure 11. As shown by Werner *et al.*, KBTBD8 knockdown in hESCs increases the amount of CNS precursors while reducing the hESC capability of differentiation into cells from the mesenchymal lineage³⁵³. Werner *et al.* hypothesized that KBTBD8-CUL3 dependent ubiquitination could mediate the interaction of RNA polymerase I with ribosome modification enzymes at specific mRNAs to regulate neural differentiation³⁵³. Hypothetically, neural differentiation could also be modulated by *BTBD8* interaction with CUL3. This would explain the differences observed in the neuronal cell-types generated from *BTBD8*-mutant and wild type hiPSCs (Figure 66, Figure 67, Figure 68, Figure 69, Figure 78), although it remains to be explored.

The proteomic analysis also revealed alterations in two neuropilins: NRP1 and NRP2. NRP1 is a cell-surface receptor for vascular endothelial growth factor (VEGF) and semaphorin (SEMA3A) involved in axon guidance, cell migration and apoptosis. NRP2 is thought to play a similar role, but in different neuronal populations. As indicated by Danelon *et al.*, alterations in NRP1/SEMA3A can lead to alterations in neuro-morphogenesis, affecting dendrite arborization and axon guidance³⁵⁴. In addition, alterations in NRP2-mediated signalling have been associated with alterations in synaptic maturation and GABAergic migration in ASD cases, as explored by Eisenberg *et al.*³⁵⁵. Furthermore, Khmour *et al.* described spine density and synchronization alterations associated with NRP2 alterations³⁵⁶. In Figure 80E-F can be observed that NRP1 is only detected in *BTBD8*-mutant cultures at both time points, while NRP2 is only present in day 60, both being absent from the wild type. The abnormal activity of the neuropilin/semaphorin pathway could be involved in the morphological alterations inferred from the proteomic data.

Finally, *BTBD8*-mutant cultures also showed alterations in the reelin mediated pathways, which are frequently suggested as putative related to ASD cases as previously mentioned in the introduction. Reelin, an extracellular matrix glycoprotein, is involved in neural migration and cortical layer establishment³⁸. Thus, the absence of CTIP2/BCL11B observed in *BTBD8*-mutant 2D cultures (Figure 78H) could be associated reelin dysfunctions mediated by *BTBD8*.

Nevertheless, positive cells for the cortical markers CTIP2/BCL11B and SATB2 were found in BTBD8-mutant brain organoids at the time points analysed (Figure 69). A delayed differentiation of cortical neurons could explain this observation, since only two time points were analysed. If that hypothesis was correct, CTIP2/BCL11B should be detected in 2D cultures differentiated for a longer period. Additionally, reelin also interacts with SNARE proteins regulating neurotransmitter release³⁵⁷, so alterations on the pathway could also lead to alterations in neuronal excitability.

Overall, more experiments are imperative to determine the implications of the proteomic findings presented in this thesis. As mentioned, an RNA-seq analysis is currently under development to study the differences between wild-type and *BTBD8*-mutant hiPSCs. Also, it would be important to conduct electrophysiology studies to assess whether the observed alterations have consequences in neural function and signal transmission. Finally, the characterization of the synaptic morphology, using techniques such as electron microscopy, could be key to confirm the role of BTBD8 in human neurons.

5.2. GENERATION AND CHARACTERIZATION OF AN *IN VIVO* MODEL TO STUDY THE FUNCTION OF *FOSL2*.

The second objective of this thesis was to develop a zebrafish knockdown model to study the role in neurodevelopment of the gene *fosl2*, which encodes a protein of the Fos family. In vertebrates, members of the Fos family are recruited into AP-1 transcription factors, through which they regulate gene expression in response to diverse environmental signals.

As stated in the introduction, most studies concerning *FOSL2* function have been focused on its implication in cancer and inflammation, but little it is known about its role during embryogenesis and neurodevelopment.

In 2022, Cospain *et al.*²²¹ described for the first time a human phenotype associated with mutations in *FOSL2*. The authors identified heterozygous truncating mutations in the last exon of *FOSL2* in individuals that shared a common phenotype characterized by developmental delay and ASD, along with defects in enamel and skull formation. Through *in vitro* assays in HEK293 cells, they identified that mutations in the last exon made *FOSL2* mRNA escape non-sense mediated decay and thus stabilized AP-1 complexes²²¹. However, the functional impact of *FOSL2* dysregulation was not further studied.

Traditionally, mouse models have been the preferred option to conduct *in vivo* assays to study gene function, but these models have limitations specially for the study of developmental processes. As commented on the introduction, zebrafish is a more suitable model to address developmental alterations since embryos develop externally and are transparent. In addition, Kubra *et al.* explored zebrafish potential to study *fos* genes³⁰⁰. In mammals, four members of the family have been identified (*C-FOS*, *FOSB*, *FOSL1* and *FOSL2*), while six are present in zebrafish, as two genes have been duplicated during evolution (*fosaa/fosab*, *foslla/fosllb*, *fosb*, and *fosl2*)³⁰⁰. Notably, *FOSL2* has only one orthologous gene in zebrafish, known as *fosl2*, which simplifies the study of its function. Kubra *et al.* also studied the expression pattern of *fos* genes in zebrafish embryos and adults, with *fosl2* having the highest expression in brain, suggesting an important role in neurodevelopmental regulation³⁰⁰. To date, *FOSL2* role in brain development has not been further explored using neither *in vivo* nor *in vitro* models. In addition, zebrafish *fosl2* has only been studied in the context of cardiac differentiation by Jahangiri *et al.*²³⁸.

Studying the role of *fosl2* in neurodevelopment can provide fundamental insights into the molecular and cellular processes that shape the brain. This knowledge not only has the potential

of improving our understanding of neurological health and disease, but also to develop novel therapeutic interventions. In this context, the model generated in this thesis is, as far as we are aware, the first model generated to specifically study *fosl2* role in neurodevelopment.

5.2.1. Phenotypic alterations in *fosl2* morphant zebrafish.

At 48 hpf, the analysis of zebrafish phenotypes revealed diminutions in body length, cephalic size, and ocular size among those treated with MO-*fosl2* (Figure 84). Additionally, there were no significant differences observed between the two treated groups: MO-*fosl2* alone and MO-*fosl2* in combination with MO-*p53*. Moreover, the head/body ratio is increased, and the eye/head ratio is reduced in morphant individuals. Altogether, these results suggest that the development of morphant embryos is delayed in comparison to control individuals of the same age^{148,189–193,358}.

As mentioned earlier, zebrafish's *fosl2* has only been studied by Jahangiri et al²³⁸. In their study, cephalic alterations are visible in *fosl2* morphant embryo's representative images. However, the authors did not address this altered phenotype since their work was focused on cardiac development. In the same publication, a KO zebrafish model was generated for *fosl2*. Regarding this model, the authors mentioned that mutant embryos were phenotypically similar to wild-type and that they reached adulthood being fertile. The absence of reported neurodevelopmental alterations in the *fosl2* model by Jahangiri et al. raises the possibility that these changes may have been overlooked. Data presented in this thesis show that *fosl2* phenotypical alterations could be easily missed at first sight (Figure 83), thus highlighting the importance of evaluating zebrafish phenotypes in a precise way. Here, phenotypes were measured in the photographed embryos taking common reference points across all images, thus conducting an objective evaluation of differences across experimental conditions (Figure 84, Figure 85).

The high mortality rate of the *fosl2* morphant embryos (Table 32) observed in this thesis was not reported in the publication by Jahangiri et al²³⁸. However, in *Mus musculus*, *Fosl2* KO pups die shortly after birth, an indicative of the relevance of this gene during development. Taking into account the lethal phenotype displayed by knockout pups, *Fosl2* has been mainly studied in mice through conditional KO strains^{359,360}. Conditional KO strains have been mainly generated to study *Fosl2* implication in inflammation and immunity^{240,361,362}, but none have been used yet to study its function during development of the nervous system.

Fosl2 has also been studied in the context of bone and cartilage development in mouse models^{236,359,360,363}, but no reports about *fosl2* implication in bone and cartilage formation have been published in zebrafish yet. Notably, the conditional KO of *Fosl2* in mice chondrocytes induces a similar phenotype to the one observed in the complete KO mice, characterized by growth retardation and premature lethality^{359,360}. Cospain et al. also reported alterations in the skull and enamel formation in *FOSL2*-associated neurodevelopmental disorder²²¹. For this reason, cartilage development was studied in *fosl2* morphant embryos at 120 hpf (Figure 86). MO-*fosl2* + MO-*p53* individuals showed a reduction in the palatoquadrate length in comparison to MO-*fosl2* zebrafish, but no other alterations were observed across conditions.

The presence of behavioural alterations was analysed at 120 hpf. The outcomes of this experiment (Figure 87, Figure 88) revealed that highest levels of anxiety-like responses are detected in zebrafish treated with MO-*fosl2* + MO-*p53*. In addition, no differences were found when comparing the control and MO-*fosl2* groups when the cycles were individually analysed.

The observed outcomes of the cartilage and behavioural assays could be explained by a potential lack of involvement of *fosl2* in those processes in *D. rerio*. However, considering the

established role of *fosl2* in bone and cartilage formation in mice and humans and the evolutionary conservation of *fosl2*, this scenario appears less plausible. An alternative explanation could be linked to the characteristics of the zebrafish model created for the study. As outlined in the introduction, the efficacy of morpholino knockdown lasts for a limited period, typically 4-5 days depending on the specific morpholino variant used. For this reason, it is possible that at 120 hpf the effects of MO-*fosl2* start to decay, while MO-p53 remains active. This temporal distinction could account for the observed differences between experimental conditions, providing a plausible reason for the discrepancies noted in the studies conducted at 120 hpf.

The outcomes of the phenotypic analyses conducted in this thesis support the suitability of zebrafish as a model to study *fosl2* function in development, as previously proposed by Kubra et al³⁰⁰. Furthermore, the generated morphant model recapitulates the growth retardation phenotype associated with *FOSL2* mutations in humans²²¹. While the primary objective of the *fosl2* knockdown model was not to precisely mimic the human neurodevelopmental condition, the evident alterations in neurodevelopmental processes induced by *fosl2* knockdown serve as a promising indication of its potential relevance to such conditions in future studies.

Future directions in the study of *fosl2*'s role in neurodevelopment would include the generation of a *fosl2* zebrafish mutant line to confirm morphant phenotypes, as several studies have found discordances between models^{148,158–160}. The creation of a *fosl2* mutant line, particularly with truncations in the last exon, would also be of particular interest. This approach holds the promise of faithfully recapitulating the neurodevelopmental phenotype associated with mutations in *FOSL2*, as detailed by Cospain et al²²¹, while also allowing the analysis of phenotypic and molecular alterations during development in an accessible way.

5.3. GENERATION AND CHARACTERIZATION OF *IN VITRO* AND *IN VIVO* MODELS TO STUDY THE FUNCTION OF *TAOK1*.

The third and final objective of this thesis was to create suitable models for scrutinizing the role of *TAOK1* in neurodevelopment. This involved the generation of knockout models both *in vitro*, utilizing hiPSCs, and *in vivo*, through the development of a zebrafish mutant line.

In contrast to the other two genes examined in this thesis, there is compelling evidence supporting the association between neurodevelopmental alterations and *TAOK1* dysfunction. As outlined in the introduction, multiple case reports detail clinical features observed in patients with *TAOK1* mutations^{246,248,253,273}. Despite this clinical insight, the molecular mechanisms underlying *TAOK1* dysfunction have received limited attention in the past.

Emphasizing the significance of this research gap, Huang *et al.* highlighted that *TAOK1* remains among the 30 % of human kinases that are understudied³⁶⁴. This is noteworthy, considering kinases are druggable targets with high evolutionary conservation, facilitating their investigation using *in vivo* models^{245,247,364}.

Drosophila melanogaster has been the most used *in vivo* model to study TAOs proteins, although this animal model only has one representative gene of the family, TAO-1, which limits the extrapolation of the findings to humans³⁶⁵. Conditional knockout mouse models had been developed to explore the role of *Taok1* in inflammation and immunity^{257–259}. However, until September 2023, there were no reports available detailing the generation of *Taok1* mouse models specifically designed for the study of neurodevelopmental conditions. Addressing this gap, Wang *et al.* established a haploinsufficiency *Taok1* mouse model to investigate its implications in ASD development³⁶⁶.

Expanding this research gap, the lack of *in vivo* models to study *TAOK1*'s role in the nervous system extends to *Danio rerio*. Despite the advantages of using this animal model for neuroscience research, it has not been yet used to study *taok1*, the orthologous gene of *TAOK1* in this organism.

In vitro models for investigating the molecular function of *TAOK1* are more prevalent. Particularly noteworthy are the studies conducted by van Woerden *et al.* and Beeman *et al.*, in which they replicated ASD-associated mutations in *TAOK1* within rodent primary neuronal cultures^{248,256}. Nevertheless, suitable human *in vitro* models to explore *TAOK1* function have not been generated until now.

The use of hiPSCs and zebrafish models represents a pivotal approach for the study of *TAOK1* function. On one side, hiPSCs offer a unique platform to model human-specific aspects of *TAOK1* function in a controlled *in vitro* environment. Moreover, these cells can differentiate into various cell types relevant not only to neurodevelopment, but also to other conditions associated with *TAOK1* dysfunction. On the other side, zebrafish offers numerous advantages in studying *TAOK1* function in a complex scenario, while allowing an easier genetic manipulation and simplified analysis of early neurodevelopmental phenotypes compared to other *in vivo* usually applied models.

Integrating findings from hiPSCs and zebrafish models could enhance our understanding of *TAOK1* role in neurodevelopment, while also contributing to the establishment of suitable platforms for testing therapeutic strategies suitable for treating neurodevelopmental conditions associated with *TAOK1* dysfunction.

5.3.1. Evaluation of the functional consequences of *TAOK1* stop gain mutations in brain organoids.

Findings from this thesis indicate that human brain organoids derived from the hiPSCs line Ctrl.BS-PBiPS37-SvF4-1-*TAOK1-i2* exhibited larger sizes than their isogenic controls, with differences being statistically significant between days 18 and 27 (Figure 112). The general association between macrocephaly and ASD has already been addressed in this discussion (section 5.1.2)^{35,303-309}, when the phenotype of *BTBD8*-mutant organoids was commented. Contrary to *BTBD8*, clinical reports are available for ASD-cases with *TAOK1* mutations, in many of which the head size was altered. Xie *et al.* identified a deletion in chromosome 17, that included *TAOK1*, in a patient diagnosed with developmental delay accompanied by microcephaly²⁷⁰. However, this is the only report that associates *TAOK1* dysfunction with a diminution in brain size. In fact, multiple authors described macrocephaly as one of the most common features of individuals with mutations in *TAOK1*^{246,248,253,271,367}. Furthermore, the phenotype of brain overgrowth was observed in several animal models, including *Mus musculus* (*Taok1*^{+/-} haploinsufficiency)³⁶⁶ and *Drosophila melanogaster* (*Tao-1* knockdown)³⁶⁵.

Unfortunately, phenotypical data is not complete for patients with *TAOK1* mutations in our cohort of ASD cases, so it is unclear whether they have macrocephaly. This shows that is essential to potentiate the crosstalk among clinicians and genetic experts. Improving the phenotypic characterization of patients would be extremely helpful to determine if the generated models can fully recapitulate the alterations observed in the clinical practice, which is necessary to advance towards a better diagnosis and personalized medicine.

Regarding the transcriptomic approach, the preliminary analysis of the scRNA-seq data showed alterations in the Cajal-Retzius cell population between Ctrl.BS-PBiPS37-SvF4-1-*TAOK1-i2* brain organoids and their isogenic controls. In wild-type brain organoids, Cajal-Retzius cells are more abundant in day 62 samples, although a small fraction of cells is already

present at day 48 (Figure 119). Meanwhile, this specific cell population is almost absent in Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1*-i2 brain organoids, even after 62 days of differentiation (Figure 120, Figure 122, Figure 123). This observation could be indicative of several alterations: a) an impairment in the differentiation of Cajal-Retzius cells in *TAOK1*-mutant brain organoids, b) a delayed differentiation of Cajal-Retzius cells in *TAOK1*-mutant brain organoids, or c) an impairment in the expression of *RELN*, the main marker of Cajal-Retzius cells. Cajal-Retzius cells, named after Santiago Ramón y Cajal and Gustaf Retzius, are a subtype of neurons that appear transiently on the brain cortex surface during development³⁶⁸. This cell population plays an important role in the spatial organization of the cells within the cerebral cortex, especially in mammals due to its complexity³⁶⁸. Through the release of Reelin during early stages of brain development, Cajal-Retzius cells have been implicated in cortical lamination and cell migration³⁶⁸⁻³⁷¹. Alterations in the Reelin pathway have been extensively associated with neuropsychiatric disorders, including ASD³⁶⁸⁻³⁷⁰. The association between *RELN* and ASD was first described by Fatemi *et al.*, who found a 44 % decrease of *RELN* expression in the cerebellum of ASD cases³⁷². On the contrary, Camacho *et al.* analyzed the density of reelin-expressing neurons in ASD postmortem samples, finding no significant differences between the brain cortex of ASD cases and controls. Thus, these authors concluded that the reduction of Reelin could be due to changes in protein processing rather than alterations in the number of neurons that synthesize it³⁷⁰. It is important to mention that Camacho *et al.* used samples obtained from adults in their study, and thus most reelin-expressing neurons are not Cajal-Retzius cells. During early postnatal stages, most part of Cajal-Retzius cells suffer apoptosis, and only a small portion persist into adulthood. Thus, in adult brains, Reelin is mainly synthesized by GABAergic interneurons^{368,371}. Taking this into account, using human postmortem brain samples to study the impact of Reelin dysregulation in neurodevelopmental disorders is limited. Future analyses are needed to explore the impact of *TAOK1* mutations in the formation of Cajal-Retzius cells and Reelin synthesis, since those connections have not been explored yet in the literature. The thorough analysis of scRNA-seq data will shed light on this topic for its biological interpretation and new hypothesis to be generated.

Moreover, scRNA-seq data analysis also showed that Ctrl.BS-PBiPS37-Sv4F-1 brain organoids differentiated from 62 days presented a cluster of cells that overexpressed ECM synthesis genes, most of them involved in collagen synthesis (Figure 123). Notably, Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1*-i2 cells were almost absent from this cluster. Alterations in ECM components have been previously found in other models of ASD, including *CHD8*^{+/-} brain organoids¹²⁴ and rats treated with valproic acid³⁷³. Remarkably, alterations in *COL3A1* could impact neuronal migration and thus, altered the correct formation of cortical layers^{373,374}, which was also suggested by the alterations in the Cajal-Retzius population previously discussed.

Consistent with our findings, impaired neuronal migration has also been identified in two mouse models: van Woerden *et al.* detected an abnormal migration in *Taok1* deficient neural progenitors, while Wang *et al.* made a similar observation in their haploinsufficiency mouse model (*Taok1*^{+/-})^{248,366}. Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1*-i2 brain organoids seem to successfully recapitulate alterations previously linked to *TAOK1* neurodevelopmental alterations, while also allowing us to explore their molecular basis in a human context.

Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1*-i2 brain organoids also show a reduced number of cells within the *NEUROD4* expressing cell cluster in comparison to the WT controls (Figure 122, Figure 123). As reviewed by Tutukova *et al.*, *NEUROD4* is mainly expressed during development in the ventricular zone and its role seems associated with neuronal differentiation. Interestingly, Hardwick *et al.*, reported that *Neurod4* activity strongly depends on its phosphorylation status in serine/threonine-proline sites and that a phospho-mutant version of

this protein promotes neuronal differentiation³⁷⁵. Regarding ASD pathogenesis, Sood *et al.* identified an up-regulation of *NEUROD4* associated with mutations in *CHD8*, a chromatin remodeler often associated with the disorder³⁷⁶. Despite no reports of *TAOK1* being implicated in Neurod4 regulation, the preliminary data obtained in this thesis suggest a potential connection between them. Future phospho-proteomic analysis of Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1*-i2 brain organoids could help exploring this possibility.

Studies conducted in *D. melanogaster* and *C. elegans* had identified orthologous genes of *TAOK1* (Tao-1 and kin-18, respectively) as regulators of neuronal polarity and differentiation^{253,365,377}, a function that seems preserved in mammalian cells²⁵². For this reason, alterations were expected in the radial glia and neuronal populations derived from *TAOK1*-mutant brain organoids in comparison to WT controls.

As anticipated by Uzquiano *et al.*²⁹⁷, aRG emerged as the predominant cell type identified in our human brain organoids at both 48 and 62 days of differentiation (Figure 119, Figure 120). Furthermore, the expected trend involves a decline in the number of aRG cells over time, aligning with the heightened maturity of the brain organoids. While this developmental trajectory was evident in WT brain organoids, it appears to progress at a slower pace in the organoids derived from the *TAOK1*-mutant hiPSCs line. Additionally, differences were also observed in the oRG populations, which seem to cluster independently according to their genotype (Figure 123). According to previous reports, *TAOK1* deficiency could impact neuronal progenitor mitosis due to altered microtubule dynamics through the Par-1/Tau pathway²⁵². In addition, studies conducted in *D. melanogaster* identified the Hippo pathway as a key regulator of neuroblast proliferation, which is controlled by Tao-1^{253,365}. However, it is necessary to note that *Drosophila* has a sole representative gene (*Tao1*) for the TAOK family, whereas humans possess three (*TAOK1*, 2 and 3). This evolutionary divergence makes it challenging to directly extrapolate findings from *Tao1* knockdown experiments in *Drosophila* to elucidate the role of *TAOK1* in mammalian systems. Future thorough examination of the expression signatures and differentially expressed genes in RG populations is needed to identify the molecular mechanisms behind the observed alterations.

Findings from this thesis also suggest potential alterations in the differentiation of astroglia. As illustrated in Figure 122 and Figure 123, cell clusters expressing astroglia markers (*GJAI*, *SL00B*, *SPARC*) primarily consist of cells derived from Ctrl.BS-PBiPS37-Sv4F-1 brain organoids. Astroglia are essential for correct neuronal functioning, regulating processes that include from neuronal migration to synaptogenesis^{378,379}. Taking this into account, it is not surprising that astrocyte dysfunctions have been identified in several neurodevelopmental disorders. In the context of ASD, several studies have identified reduced number of astrocytes in ASD models. In addition, excitatory/inhibitory imbalances could also be associated with astrocyte alterations^{378,379}. Nevertheless, research in this field is limited, and little is known about the direct association between astroglia and ASD. The role of *TAOK1* in astrocyte differentiation and function has not been described yet, but some authors suggest its potential implication in the regulation of microtubule dynamics³⁸⁰.

Regarding the presence of excitatory neurons, deep-layer projection neuron clusters were identified both in WT and *TAOK1*-mutant samples (Figure 122). Interestingly, in day 62 samples, two clusters of CFuPN neurons were identified, comprising almost exclusively *TAOK1*-mutant cells (Figure 123).

Differences between WT and *TAOK1*-mutant brain organoids were also detected in clusters expressing GABAergic identity markers (*GAD2*). Notably, clusters of subcortical GABAergic interneurons were mainly composed of cells derived from Ctrl.BS-PBiPS37-Sv4F-1 brain organoids, while immature GABAergic cortical interneurons were mainly composed of

TAOK1-mutant cells (Figure 123). These observations indicate an aberrant neuronal differentiation process in Ctrl.BS-PBiPS37-SvF4-1-*TAOK1-i2*, wherein neural differentiation appears to occur at a faster-than-expected pace.

In the context of cortical neuron differentiation, Wang *et al.* observed a reduced density of neurons in the upper layer cortex of *Taok1*^{+/-} adult mice³⁶⁶. However, the primary focus of their investigation relied on the effects of *Taok1* haploinsufficiency in the subcortical dorsal raphe nucleus, a structure not developed in the brain organoids generated in this thesis^{119,297}. Interestingly, Paulsen *et al.* also identified asynchronous development deep-layer excitatory projection neurons and GABAergic neurons in brain organoids with mutations in three ASD-associated genes (*CHD8*, *SUV420H1*, and *ARID1B*)¹²³. Despite these intriguing findings, the precise molecular mechanisms responsible for these shared phenotypes remain elusive. Work in progress is being made towards the exploration of our scRNA-seq data, with an emphasis on identifying gene expression patterns, cellular interactions, and signalling pathways. By doing so, we aim to improve our understanding of common neurodevelopmental alterations shared across ASD risk genes.

Impairments in neurite formation have also been associated with *TAOK1* mutations^{252,253}. To investigate the underlying mechanisms, Beeman *et al.* conducted a study using primary hippocampal neurons from rats²⁵⁶. They examined the impact of various missense mutations, generating catalytically inactive forms of TAOK1, all previously associated with neurodevelopmental disorders. According to their findings, TAOK1 acts as a plasma remodelling kinase, and activity that is regulated by autophosphorylation. Thus, missense mutations could lead to impairments in neuronal morphogenesis, migration, and maturation due to alterations in the formation of the dendritic arbour. Although morphological alterations have not been yet examined in our model, it has the potential of unravelling the effects stop-gain mutations in the morphogenesis neurons and glia cells, as the knockout model was generated in hiPSCs.

Undoubtedly, the obtained scRNA-seq data requires thorough exploration, and the verification of cluster characterization is fundamental. Nonetheless, the preliminary results appear promising, unveiling new research avenues to investigate the role of *TAOK1* in neurodevelopment. In addition, the edited cell line generated in this thesis, Ctrl.BS-PBiPS37-SvF4-1-*TAOK1-i2*, holds potential for advancing our understanding of *TAOK1* function through in-depth phospho-proteomic and electrophysiological analyses.

As a final remark, it is noteworthy that Wang *et al.* successfully alleviated the behavioural alterations observed in their *Taok1* haploinsufficiency mouse model. This amelioration was accomplished through the targeted injection of AAVs carrying the wild-type kinase into the dorsal raphe nucleus of *Taok1*^{+/-} mice³⁶⁶. Considering these promising results, our laboratory is currently exploring the effectiveness of various delivery vectors, such as lipid nanoparticles and AAV vectors, in brain organoids. Once the optimization is over, the knockout model developed in this thesis could be useful to test whether re-introducing WT *TAOK1* could reverse the observed phenotypical alterations and its potential as a gene therapy strategy.

5.3.2. Phenotypic alterations in *taok1* mutant zebrafish.

Only minor morphological alterations were detected during the phenotypic characterization of *taok1* mutant zebrafish. On one side, at 48 hpf, the reduction in the number of *taok1* wild type alleles correlated with a slight decrease in cephalic size, body length, and ocular size. However, due to the dispersion of the data, most of these observations were non-significant, except for the reduction of cephalic size in heterozygous *taok1*^{415/419} embryos. Nevertheless,

this significant reduction in size was no longer apparent in 96 hpf larvae. At 96 hpf, individuals with the heterozygous *taokl*^{415/419} genotype exhibited diminished ocular size, with no other discernible morphological alterations. These observations stand in contrast to clinical features of individuals with *TAOK1* mutations, which often include macrocephaly^{246,248,253,271,367}. In addition, previously generated *in vivo* models in *Mus musculus* (*Taok1*^{+/-})³⁶⁶ and *Drosophila melanogaster* (*Tao-1* knockdown)³⁶⁵ also share this morphological feature. The absence of macrocephaly in *taokl* mutant zebrafish could be associated with biological differences between species but could also be a phenotype apparent later in development. To thoroughly validate the absence of morphological alterations in the zebrafish mutant model, it is imperative to broaden the morphological characterization beyond the current scope. This would involve incorporating additional time points across larvae and adult stages.

Interestingly, in their publication from September 2023, Wang *et al.* noted that *Taok1*^{-/-} mutations were lethal in *Mus musculus*, leading them to an haploinsufficiency model (*Taok1*^{+/-}) for their experiments³⁶⁶. A similar observation was made by Dulovic *et al.* in *Drosophila melanogaster*, where *Tao-1* knockdown was associated with a premature death of the individuals²⁵³. In this thesis, homozygous individuals with mutations in *taokl* were successfully obtained, although they were euthanized after the 48 hpf analysis. Notably, in the F1 generation, heterozygous individuals (*taokl*^{415/419}) demonstrated an unexpectedly early mortality, leaving only one surviving adult male. This outcome has constrained the scope of the analysis conducted as part of this thesis, as previously addressed in the results section. Future analyses will elucidate whether *taokl* homozygous mutations indeed result in premature death and how heterozygous mutations impact the life expectancy of *Danio rerio*.

Behavioural analysis conducted in this thesis revealed that anxiety-like responses after illumination changes were very similar between *taokl*^{415/419} individuals and controls at 96 hpf. These results align with observations conducted by Wang *et al.*, who reported a similar behaviour in *Taok1*^{+/-} mice³⁶⁶. Results from this thesis also indicate that 96 hpf *taokl*^{415/419} individuals are overall less active than WT counterparts under light conditions. This observation contrast with previous reports in *Mus musculus*, on which *Taok1*^{+/-} individuals typically demonstrate heightened activity levels³⁶⁶. However, it is important to notice that the experiments in mice were performed with adults, which can impact the observed phenotypes. To address this, future behavioural analyses will also encompass adult zebrafish, providing a deeper examination of *taokl*^{415/419} behavioural impact.

To the best of our knowledge, no other morphant or mutant models for *taokl* have been developed in zebrafish, making this study the first characterization of the morphological and behavioural alterations associated with *taokl* modifications in this animal model.

Undoubtedly, the results presented in this thesis are a preliminary characterization of *taokl* mutant phenotype in zebrafish, but more works needs to be done to fully characterize the line. The limited number of mutant individuals impeded the knockout verification, which is essential for the correct interpretation of the results. Work in progress is being made towards addressing this issue.

Furthermore, it is crucial to contemplate the possibility of compensatory effects attributable to other members of the *TAOK* family. As mentioned in the introduction, *ENSDARG00000098304* is a potential paralogous sequence of *taokl*. Therefore, it is imperative to ascertain whether its expression is altered in the mutant model developed in this thesis. Additionally, orthologous sequences of *TAOK2* and *TAOK3* could also play a role in this context and warrant investigation.

After conducting these validations, it would be compelling to broaden the phenotypic characterization of the mutant model, incorporating additional time points in the analysis,

particularly focusing on adults. In addition, exploring transcriptomic and proteomic profiles in the mutant model could prove beneficial in unravelling the role of *taok1* in neurodevelopment and assessing its conservation across vertebrates. Such insights would be crucial to ascertain the suitability of zebrafish as a preclinical *in vivo* model for prospective therapeutic interventions targeting *TAOK1*.

5.4. GENERAL CONSIDERATIONS REGARDING THE *IN VITRO* MODELS GENERATED IN THIS THESIS.

All the *in vitro* models generated in this thesis derive from the same control hiPSCs line, Ctrl.BS-PBiPS37-Sv4F-1. This cell line was edited with the CRISPR-Cas9 system to generate knockout mutations in two genes of interest, *BTBD8* and *TAOK1*, independently. This approach has the advantage of maintaining the genetic background of the control cell line, thus allowing us to study the functional consequences of altering our genes of interest in a controlled way. However, considering the complexity of ASD, it would also be interesting to analyse if the same phenotypes arise when the mutations are introduced into other WT cell lines, to rule out compounding effects due to the genetic background of our cell line.

In this thesis, the selected approach for studying genetic variations potentially linked to ASD involved the knockout of genes of interest. By doing so, the main objective was to improve our knowledge about gene function and their association with neurodevelopmental alterations, as these connections had not been deeply studied before. However, future approaches might involve the reprogramming of somatic cells from patients into hiPSCs to study the specific effects of their mutations preserving the genomic background, followed by the application of precise genome editing techniques to revert them into the wild type sequence.

Edited hiPSCs lines were generated using a transfection-based delivery of the CRISPR-Cas9 components. This method has the advantages of not requiring specific equipment and being non-integrative, so the components of the CRISPR-Cas9 system are eliminated in a short period of time²⁷⁹. The characterization of the edited cell lines was extended, including pluripotency assessment, off-target evaluation, and detection of major genomic aberrations. However, the techniques applied have several limitations that need to be addressed.

The assessment of unintended genome editing events involved the analysis of off-target regions predicted *in silico*. Nevertheless, in the future, NGS is intended to be the chosen method for these analyses as it is a more comprehensive and unbiased alternative for identifying unintended editing events³⁸¹.

Another limitation associated with the use of hiPSCs is their genomic and epigenomic instability, which increases after being maintained in culture for extended periods of time or being subjected to stressful event, such as genome editing^{303,382,383}. In the present work, the presence of genomic alterations was addressed with a karyotype characterization. Karyotype analysis allows the detection of major genomic rearrangements common in hiPSCs, such as trisomies (more common in chromosomes 1, 8, 12, 17, and X), and large duplications (such as 20q11.2)^{382,383}. As exposed earlier in the results, these major alterations were successfully ruled out in both edited hiPSC lines generated throughout this thesis (Figure 52, Figure 101). While karyotype analysis offers a cost-effective and single-cell resolution, it has limitations, particularly in detecting small genetic alterations. Therefore, additional measures, such as SNP arrays, WGS, or employing several clonal lines in experiments would be required to address this issue. Nevertheless, those approaches come with a high associated cost for processing few samples, which can be a limiting factor for their application^{303,381,382}. Epigenetic alterations were not characterized in the developed hiPSCs models as a part of this thesis. Future analysis

of the epigenetic landscape could involve DNA methylation profiling and chromatin immunoprecipitation sequencing (ChIP-seq) to further characterize the edited hiPSCs³⁸³.

Apart from the inherent limitations of using hiPSCs lines edited with the CRISPR-Cas9 technology, the differentiation protocols that were used to differentiate them has also a major impact on the research's outcomes.

The differentiation of cortical neurons (2D) from hiPSCs was conducted following the indication of Shi *et al.* with minor modifications¹⁰⁵. This protocol, published in 2012, is based on dual SMAD inhibition and yield a fair amount of deep-layer cortical neurons (CTIP2/BCL11B⁺). However, in our hand, its reproducibility was limited, which impacted the number of successful differentiations performed. As reviewed by Yan *et al.*, there are more recent advancements in dual SMAD inhibition methods, yet challenges persist in achieving high levels of maturation in the obtained neurons^{304,384}.

The differentiation of hiPSCs into more complex and representative models of the developmental brain was also addressed in this thesis. Human dorsal forebrain organoids were generated following the protocol by Velasco *et al.* with minor modifications^{119,286}. In our hands, this protocol was highly reproducible, and the obtained scRNAseq results align with the expected diversity of cell populations referenced by the authors, which resembles foetal development. Briefly, neuroepithelial cells proliferate and generate aRG, IP, and oRG populations. The differentiation of these progenitors originates different cell types, mainly excitatory neurons in the dorsal region and interneurons in the ventral region^{119,122,297}. Regrettably, electrophysiology assays were not conducted as part of this thesis to identify alterations in neuronal function. Despite this limitation, the protocols for generating neuronal cells, both in 2D and 3D, were optimized. It is anticipated that these optimizations will pave the way for future electrophysiology assays to be performed in the near future.

An alternative strategy consists in the combination of spheroids and/or organoids from different origins to generate assembloids, which can be used to study complex circuits dynamics and neuronal migration between regions^{304,385}. Miura *et al.* have successfully applied this novel approach to generate cortical-striatal assembloids, a very interesting model for the identification of alterations associated with ASD since this pathway has been associated with the disorder³⁸⁶. This protocol was also tested as a part of this thesis, but, unfortunately, it was not possible to optimize it on time for its inclusion in the present work. Another exciting possibility involves the combination of neuronal and non-neuronal spheroids to study the intricate mechanisms leading to ASD pathology, such as the impact of the brain-gut axis or the role of immune cells in the development of the disorder.

The most recent advance in the field of modelling development in vitro consist in the generation synthetic embryos, also known as embryoids. Regarding this novel approach, Amin *et al.* were able to reproduce the first 8 days of mouse development using mESCs³⁸⁷, although the protocol is still to be improved.

While current protocols are not able to reproduce the complexity of human brain, the continuous advancements in the field could lead to a future scenario where concerns about rudimentary forms of consciousness arise. The rapid progress in neurodevelopmental modelling underscores the importance of robust regulatory frameworks that take these possibilities into account. This foresight is crucial not only for the ethical conduct of research but also for establishing comprehensive public guidelines that conscientiously regulate their use, fostering a balance between scientific progress and ethical considerations.

CONCLUSIONS

6. CONCLUSIONS

1. Cellular models based on hiPSCs are suitable for the development of biotechnological strategies, such as genome editing systems based on CRISPR. The combination of both enables the study and functional validation of genomic variations in pathologies using an *in vitro* model capable of recapitulating human developmental processes. Specifically, in the case of ASD, these models facilitate progress in understanding the molecular bases of the disorder, as well as the generation of advanced models for exploring novel therapeutic strategies to ameliorate ASD-associated symptoms.
2. Two *in vitro* models based on hiPSCs have been successfully generated for the functional characterization of *BTBD8* and *TAOK1*, two genes with mutations associated with ASD development. Modified hiPSC lines were created through transfection-based delivery of CRISPR-Cas9 components into the control line Ctrl.BS-PBiPS37-Sv4F-1, thus allowing transient activity of the editing system and the maintenance of the control genetic background in the edited cell lines. In addition, edited hiPSCs cells were thoroughly characterized, including pluripotency assessment, off-target evaluation, and detection of major genomic aberrations, thus ensuring the reliability of the results obtained in this thesis.
3. Zebrafish models are suitable for the study of neurodevelopmental disorders, allowing the detection of morphological and behavioural alterations that cannot be identified using *in vitro* systems.
4. The immunohistochemical analysis of *BTBD8* models is consistent with proteomic data derived from hiPSCs differentiated neurons and paves the way for future studies regarding the role of BTBD8 in the formation and function of cells from the CNS.
5. Human brain organoids derived from the hiPSCs line Ctrl.BS-PBiPS37-SvF4-1-*BTBD8*-i2 have a significant increase in size in comparison to isogenic controls. This macrocephaly phenotype, in addition to the dysregulated apoptotic and protein degradation pathways observed in cortical neurons, suggest an imbalance between proliferation and differentiation of neural progenitors associated with *BTBD8* dysfunction.
6. Ctrl.BS-PBiPS37-SvF4-1-*BTBD8*-i2 neuronal cultures show signs of neuronal differentiation impairments, including an elevated expression of NSCs markers, a reduced expression of GABAergic neuronal markers and delayed differentiation of excitatory cortical neurons, which could lead to developmental delay and excitatory/inhibitory imbalances commonly associated with ASD.
7. Cortical neurons derived from the hiPSCs line Ctrl.BS-PBiPS37-SvF4-1-*BTBD8*-i2 show impairments in clathrin-mediated endocytosis, neurotransmitter exocytosis, vesicle trafficking, and synaptogenesis, thus suggesting alterations in the excitability of the cells as a consequence of BTBD8 dysfunction.
8. The results of the phenotypic analyses conducted in this thesis support the suitability of zebrafish as a model to study *fosl2* function in development.
9. Morphant zebrafish embryos for *fosl2* show a morphological phenotype consistent with delayed development at 48 hpf, with diminutions in body length, cephalic size, ocular size, eye/head ratio, and increased head/body ratio.

10. The high mortality rate of the *fosl2* morphant embryos suggest a potential lethal phenotype associated with the reduced expression of this gene, highlighting its importance during development across vertebrates.
11. The immunohistochemical characterization of Ctrl.BS-PBiPS37-SvF4-1-*TAOK1*-i2 brain organoids, combined with scRNAseq data, has allowed the identification of phenotypic and gene expression variation in the generated *in vitro* models.
12. The macrocephaly phenotype observed in Ctrl.BS-PBiPS37-SvF4-1-*TAOK1*-i2 brain organoids, combined with alterations observed in neural progenitor cells, suggest an imbalance between proliferation and differentiation of neural progenitors associated with *TAOK1* dysfunction.
13. Alterations in the Cajal-Retzius cell population and cells expressing ECM synthesis genes suggest impairments in cortical layer formation and neuronal migration in Ctrl.BS-PBiPS37-SvF4-1-*TAOK1*-i2 brain organoids.
14. Ctrl.BS-PBiPS37-SvF4-1-*TAOK1*-i2 brain organoids show aberrant neural differentiation affecting astroglia, excitatory and inhibitory neurons. More specifically, *TAOK1* stop gain mutations seem to accelerate the differentiation of deep layer corticofugal projection neurons and GABAergic cortical interneurons, while also impairing the differentiation of subcortical GABAergic interneurons and astroglia.
15. Morphological observations made in the zebrafish mutant model do not align with reported phenotypes of *TAOK1* mutant animal models. However, the zebrafish knockout model's development was only analysed until 48 hpf, so studies of longer duration must be conducted to confirm these observations.
16. Zebrafish *taok1*^{415/419} individuals are overall less active than WT counterparts under light conditions, which suggest a reduced locomotor activity associated with *taok1* dysfunction.
17. The proteomic and scRNA-seq data obtained in this thesis requires thorough exploration in the future. Nonetheless, the preliminary results appear promising, unveiling new research avenues to investigate the role of *BTBD8* and *TAOK1* in neurodevelopment.
18. The *in vitro* and *in vivo* models generated in this thesis are useful tools for advancing our understanding of the molecular basis of ASD, as well as for the future development of advanced therapeutic strategies to ameliorate the symptoms of these disorders.
19. Personalized medicine requires disease models that faithfully reproduce observed phenotypes, with or without a clinical diagnosis. These models complement genetic studies, as a percentage of cases remain unexplained due to findings of variation with uncertain significance that cannot contribute to a diagnosis, prognosis, or family genetic counselling. This new understanding of medicine as personalized and multidisciplinary biomedicine is only possible with a strong commitment to modelling and advanced therapies on such models becoming a reality in healthcare systems.

BIBLIOGRAPHY

7. BIBLIOGRAPHY

1. World Health Organization (WHO). International Statistical Classification of Diseases, Eleventh Revision (ICD-11). <https://icd.who.int/browse> (2022).
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (5th ed., text rev.)*. (American Psychiatric Association, 2022).
3. Zeidan, J. *et al.* Global prevalence of autism: A systematic review update. *Autism Res. Off. J. Int. Soc. Autism Res.* **15**, 778–790 (2022).
4. Morales-Hidalgo, P., Roigé-Castellví, J., Hernández-Martínez, C., Voltas, N. & Canals, J. Prevalence and Characteristics of Autism Spectrum Disorder Among Spanish School-Age Children. *J. Autism Dev. Disord.* **48**, 3176–3190 (2018).
5. Lyall, K. *et al.* The Changing Epidemiology of Autism Spectrum Disorders. *Annu. Rev. Public Health* **38**, 81–102 (2017).
6. Elsabbagh, M. Linking risk factors and outcomes in autism spectrum disorder: is there evidence for resilience? *BMJ* **368**, l6880 (2020).
7. Global Burden of Disease Study 2019 (GBD 2019). *Institute for Health Metrics and Evaluation*. <https://vizhub.healthdata.org/gbd-results>.
8. Prevalence of autistic spectrum disorder, 2019. *IHME, Global Burden of Disease*. <https://ourworldindata.org/grapher/prevalence-of-autistic-spectrum>.
9. Loomes, R., Hull, L. & Mandy, W. P. L. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **56**, 466–474 (2017).
10. R Core Team. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput. Vienna Austria* (2020).
11. Pensado-López, A., Veiga-Rúa, S., Carracedo, Á., Allegue, C. & Sánchez, L. Experimental Models to Study Autism Spectrum Disorders: hiPSCs, Rodents and Zebrafish. *Genes* **11**, 1376 (2020).
12. Modabbernia, A., Velthorst, E. & Reichenberg, A. Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. *Mol. Autism* **8**, 13 (2017).
13. Tick, B., Bolton, P., Happé, F., Rutter, M. & Rijdsdijk, F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J. Child Psychol. Psychiatry* **57**, 585–595 (2016).
14. Genovese, A. & Butler, M. G. Clinical Assessment, Genetics, and Treatment Approaches in Autism Spectrum Disorder (ASD). *Int. J. Mol. Sci.* **21**, 4726 (2020).
15. Hallmayer, J. *et al.* Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. *Arch. Gen. Psychiatry* **68**, 1095–1102 (2011).
16. Persico, A. M. & Bourgeron, T. Searching for ways out of the autism maze: genetic, epigenetic and environmental clues. *Trends Neurosci.* **29**, 349–358 (2006).
17. Kinney, D. K., Barch, D. H., Chayka, B., Napoleon, S. & Munir, K. M. Environmental Risk Factors for Autism: Do They Help Cause De Novo Genetic Mutations That Contribute to the Disorder? *Med. Hypotheses* **74**, 102–106 (2010).
18. Ornoy, A., Weinstein-Fudim, L. & Ergaz, Z. Prenatal factors associated with autism spectrum disorder (ASD). *Reprod. Toxicol. Elmsford N* **56**, 155–169 (2015).

19. Woodbury-Smith, M. & Scherer, S. W. Progress in the genetics of autism spectrum disorder. *Dev. Med. Child Neurol.* **60**, 445–451 (2018).
20. Holt, R. *et al.* Linkage and candidate gene studies of autism spectrum disorders in European populations. *Eur. J. Hum. Genet.* **18**, 1013–1019 (2010).
21. Freitag, C. M. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol. Psychiatry* **12**, 2–22 (2007).
22. Alonso-Gonzalez, A., Rodriguez-Fontenla, C. & Carracedo, A. De novo Mutations (DNMs) in Autism Spectrum Disorder (ASD): Pathway and Network Analysis. *Front. Genet.* **9**, 406 (2018).
23. Turner, T. N. *et al.* Genomic patterns of de novo mutation in simplex autism. *Cell* **171**, 710–722.e12 (2017).
24. Krupp, D. R. *et al.* Exonic Mosaic Mutations Contribute Risk for Autism Spectrum Disorder. *Am. J. Hum. Genet.* **101**, 369–390 (2017).
25. Lim, E. T. *et al.* Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. *Nat. Neurosci.* **20**, 1217–1224 (2017).
26. Alonso-Gonzalez, A. *et al.* Exploring the biological role of postzygotic and germinal de novo mutations in ASD. *Sci. Rep.* **11**, 319 (2021).
27. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568–584.e23 (2020).
28. Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
29. Yu, T. W. *et al.* Using whole-exome sequencing to identify inherited causes of autism. *Neuron* **77**, 259–273 (2013).
30. Matuszek, G. & Talebizadeh, Z. Autism genetic database (AGD): a comprehensive database including autism susceptibility gene-CNVs integrated with known noncoding RNAs and fragile sites. *BMC Med. Genet.* **10**, 102 (2009).
31. Yang, C. *et al.* AutismKB 2.0: a knowledgebase for the genetic evidence of autism spectrum disorder. *Database J. Biol. Databases Curation* **2018**, bay106 (2018).
32. Abrahams, B. S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).
33. Basu, S. N., Kollu, R. & Banerjee-Basu, S. AutDB: a gene reference resource for autism research. *Nucleic Acids Res.* **37**, D832–836 (2009).
34. Courchesne, E. & Pierce, K. Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Curr. Opin. Neurobiol.* **15**, 225–230 (2005).
35. Donovan, A. P. A. & Basson, M. A. The neuroanatomy of autism – a developmental perspective. *J. Anat.* **230**, 4–15 (2017).
36. Thabault, M. *et al.* Cerebellar and Striatal Implications in Autism Spectrum Disorders: From Clinical Observations to Animal Models. *Int. J. Mol. Sci.* **23**, 2294 (2022).
37. Ayub, R. *et al.* Thalamocortical connectivity is associated with autism symptoms in high-functioning adults with autism and typically developing adults. *Transl. Psychiatry* **11**, 1–9 (2021).
38. Scala, M. *et al.* The Pathophysiological Link Between Reelin and Autism: Overview and New Insights. *Front. Genet.* **13**, 869002 (2022).
39. Brant, B. *et al.* IQSEC2 mutation associated with epilepsy, intellectual disability, and autism results in hyperexcitability of patient-derived neurons and deficient synaptic transmission. *Mol. Psychiatry* **26**, 7498–7508 (2021).

40. Hollestein, V. *et al.* Excitatory/inhibitory imbalance in autism: the role of glutamate and GABA gene-sets in symptoms and cortical brain structure. *Transl. Psychiatry* **13**, 1–9 (2023).
41. Willsey, H. R., Willsey, A. J., Wang, B. & State, M. W. Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. *Nat. Rev. Neurosci.* **23**, 323–341 (2022).
42. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
43. Carlsson, L. H. *et al.* Coexisting disorders and problems in preschool children with autism spectrum disorders. *Sci. World J.* **2013**, 213979 (2013).
44. Salazar, F. *et al.* Co-occurring Psychiatric Disorders in Preschool and Elementary School-Aged Children with Autism Spectrum Disorder. *J. Autism Dev. Disord.* **45**, 2283–2294 (2015).
45. DeFilippis, M. & Wagner, K. D. Treatment of Autism Spectrum Disorder in Children and Adolescents. *Psychopharmacol. Bull.* **46**, 18–41 (2016).
46. Farmer, C., Thurm, A. & Grant, P. Pharmacotherapy for the core symptoms in autistic disorder: current status of the research. *Drugs* **73**, 303–314 (2013).
47. Sharma, S. R., Gonda, X. & Tarazi, F. I. Autism Spectrum Disorder: Classification, diagnosis and therapy. *Pharmacol. Ther.* **190**, 91–104 (2018).
48. Jacob, S. *et al.* Large multicenter randomized trials in autism: key insights gained from the balovaptan clinical development program. *Mol. Autism* **13**, 25 (2022).
49. Lanigan, T. M., Kopera, H. C. & Saunders, T. L. Principles of Genetic Engineering. *Genes* **11**, 291 (2020).
50. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
51. Brookhouser, N., Raman, S., Potts, C. & Brafman, David. A. May I Cut in? Gene Editing Approaches in Human Induced Pluripotent Stem Cells. *Cells* **6**, 5 (2017).
52. Khan, S. H. Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application. *Mol. Ther. - Nucleic Acids* **16**, 326–334 (2019).
53. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1156–1160 (1996).
54. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757–761 (2010).
55. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
56. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).
57. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
58. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550**, 280–284 (2017).
59. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
60. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).

61. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973–976 (2013).
62. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
63. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
64. Smith, C. *et al.* Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* **15**, 12–13 (2014).
65. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132–141 (2014).
66. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR–Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490–507 (2019).
67. Akcakaya, P. *et al.* In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature* **561**, 416–419 (2018).
68. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
69. Li, A. *et al.* Cytosine base editing systems with minimized off-target effect and molecular size. *Nat. Commun.* **13**, 4531 (2022).
70. Liang, S.-Q. *et al.* Genome-wide profiling of prime editor off-target sites in vitro and in vivo using PE-tag. *Nat. Methods* **20**, 898–907 (2023).
71. Amit, I. *et al.* CRISPECTOR provides accurate estimation of genome editing translocation and off-target activity from comparative NGS data. *Nat. Commun.* **12**, 3042 (2021).
72. Höijer, I. *et al.* CRISPR-Cas9 induces large structural variants at on-target and off-target sites in vivo that segregate across generations. *Nat. Commun.* **13**, 627 (2022).
73. Zuccaro, M. V. *et al.* Allele-Specific Chromosome Removal after Cas9 Cleavage in Human Embryos. *Cell* **183**, 1650–1664.e15 (2020).
74. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
75. Wagner, D. L. *et al.* High prevalence of *Streptococcus pyogenes* Cas9-reactive T cells within the adult human population. *Nat. Med.* **25**, 242–248 (2019).
76. Charlesworth, C. T. *et al.* Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* **25**, 249–254 (2019).
77. Burstein, D. *et al.* New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).
78. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).
79. Ihry, R. J. *et al.* p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).
80. Kim, E. *et al.* In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* **8**, 14500 (2017).
81. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139–142 (2015).
82. Finn, J. D. *et al.* A Single Administration of CRISPR/Cas9 Lipid Nanoparticles Achieves Robust and Persistent In Vivo Genome Editing. *Cell Rep.* **22**, 2227–2235 (2018).

83. Liu, J. *et al.* Fast and Efficient CRISPR/Cas9 Genome Editing In Vivo Enabled by Bio-reducible Lipid and Messenger RNA Nanoparticles. *Adv. Mater. Deerfield Beach Fla* **31**, e1902575 (2019).
84. Zakrzewski, W., Dobrzyński, M., Szymonowicz, M. & Rybak, Z. Stem cells: past, present, and future. *Stem Cell Res. Ther.* **10**, 68 (2019).
85. National Research Council (US) and Institute of Medicine (US) Committee on the Biological and Biomedical Applications of Stem Cell. Embryonic Stem Cells. in *Stem Cells and the Future of Regenerative Medicine* (National Academies Press (US), 2002).
86. Jefatura del Estado. *Ley 14/2007, de 3 de julio, de Investigación biomédica*. vol. BOE-A-2007-12945 28826–28848 (2007).
87. Ministerio de Ciencia e Innovación. *Real Decreto 1527/2010, de 15 de noviembre, por el que se regulan la Comisión de Garantías para la Donación y Utilización de Células y Tejidos Humanos y el Registro de Proyectos de Investigación*. vol. BOE-A-2010-18654 101083–101091 (2010).
88. Ministerio de Sanidad y Política Social. *Real Decreto 42/2010, de 15 de enero, por el que se regula la Comisión Nacional de Reproducción Humana Asistida*. vol. BOE-A-2010-1705 9810–9815 (2010).
89. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).
90. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
91. Chailangkarn, T., Acab, A. & Muotri, A. R. Modeling neurodevelopmental disorders using human neurons. *Curr. Opin. Neurobiol.* **22**, 785–790 (2012).
92. Prilutsky, D. *et al.* iPSC-derived neurons as a higher-throughput readout for autism: promises and pitfalls. *Trends Mol. Med.* **20**, 91–104 (2014).
93. Yu, J. *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* **318**, 1917–1920 (2007).
94. Omole, A. E. & Fakoya, A. O. J. Ten years of progress and promise of induced pluripotent stem cells: historical origins, characteristics, mechanisms, limitations, and potential applications. *PeerJ* **6**, e4370 (2018).
95. Teshigawara, R., Cho, J., Kameda, M. & Tada, T. Mechanism of human somatic reprogramming to iPS cell. *Lab. Invest.* **97**, 1152–1157 (2017).
96. Malik, N. & Rao, M. S. A review of the methods for human iPSC derivation. *Methods Mol. Biol. Clifton NJ* **997**, 23–33 (2013).
97. Bellin, M., Marchetto, M. C., Gage, F. H. & Mummery, C. L. Induced pluripotent stem cells: the new patient? *Nat. Rev. Mol. Cell Biol.* **13**, 713–726 (2012).
98. Sullivan, S. *et al.* Quality control guidelines for clinical-grade human induced pluripotent stem cell lines. *Regen. Med.* **13**, 859–866 (2018).
99. Bassett, A. R. Editing the genome of hiPSC with CRISPR/Cas9: disease models. *Mamm. Genome* **28**, 348–364 (2017).
100. Shi, Y., Inoue, H., Wu, J. C. & Yamanaka, S. Induced pluripotent stem cell technology: a decade of progress. *Nat. Rev. Drug Discov.* **16**, 115–130 (2017).
101. Ohnuki, M. & Takahashi, K. Present and future challenges of induced pluripotent stem cells. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140367 (2015).
102. Yamanaka, S. Pluripotent Stem Cell-Based Cell Therapy-Promise and Challenges. *Cell Stem Cell* **27**, 523–531 (2020).
103. Little, D., Ketteler, R., Gissen, P. & Devine, M. J. Using stem cell-derived neurons in drug screening for neurological diseases. *Neurobiol. Aging* **78**, 130–141 (2019).

104. Watson, L. M., Wong, M. M. K., Vowles, J., Cowley, S. A. & Becker, E. B. E. A Simplified Method for Generating Purkinje Cells from Human-Induced Pluripotent Stem Cells. *Cerebellum Lond. Engl.* **17**, 419–427 (2018).
105. Shi, Y., Kirwan, P. & Livesey, F. J. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. *Nat. Protoc.* **7**, 1836–1846 (2012).
106. Liu, Y. *et al.* Directed differentiation of forebrain GABA interneurons from human pluripotent stem cells. *Nat. Protoc.* **8**, 1670–1679 (2013).
107. Fedele, S. *et al.* Expansion of human midbrain floor plate progenitors from induced pluripotent stem cells increases dopaminergic neuron differentiation potential. *Sci. Rep.* **7**, 6036 (2017).
108. Speicher, A. M., Wiendl, H., Meuth, S. G. & Pawlowski, M. Generating microglia from human pluripotent stem cells: novel in vitro models for the study of neurodegeneration. *Mol. Neurodegener.* **14**, 46 (2019).
109. Canals, I. *et al.* Rapid and efficient induction of functional astrocytes from human pluripotent stem cells. *Nat. Methods* **15**, 693–696 (2018).
110. Hu, B.-Y., Du, Z.-W. & Zhang, S.-C. Differentiation of human oligodendrocytes from pluripotent stem cells. *Nat. Protoc.* **4**, 1614–1622 (2009).
111. Adhya, D. *et al.* Atypical Neurogenesis in Induced Pluripotent Stem Cells From Autistic Individuals. *Biol. Psychiatry* **89**, 486–496 (2021).
112. Deneault, E. *et al.* CNTN5-/+or EHMT2-/+human iPSC-derived neurons from individuals with autism develop hyperactive neuronal networks. *eLife* **8**, e40092 (2019).
113. Mateos-Aparicio, P., Bello, S. A. & Rodríguez-Moreno, A. Challenges in Physiological Phenotyping of hiPSC-Derived Neurons: From 2D Cultures to 3D Brain Organoids. *Front. Cell Dev. Biol.* **8**, (2020).
114. Santos, J. L. de S., Araújo, C. de A., Rocha, C. A. G., Costa-Ferro, Z. S. M. & Souza, B. S. de F. Modeling Autism Spectrum Disorders with Induced Pluripotent Stem Cell-Derived Brain Organoids. *Biomolecules* **13**, 260 (2023).
115. Paşca, S. P. *et al.* A nomenclature consensus for nervous system organoids and assembloids. *Nature* **609**, 907–910 (2022).
116. Di Lullo, E. & Kriegstein, A. R. The use of brain organoids to investigate neural development and disease. *Nat. Rev. Neurosci.* **18**, 573–584 (2017).
117. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
118. Miura, Y. *et al.* Engineering brain assembloids to interrogate human neural circuits. *protocols.io* <https://www.protocols.io/view/engineering-brain-assembloids-to-interrogate-human-bznap5ae> (2022).
119. Velasco, S. *et al.* Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019).
120. de Jong, J. O. *et al.* Cortical overgrowth in a preclinical forebrain organoid model of CNTNAP2-associated autism spectrum disorder. *Nat. Commun.* **12**, 4087 (2021).
121. Mariani, J. *et al.* FOXP1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* **162**, (2015).
122. Li, C. *et al.* Single-cell brain organoid screening identifies developmental defects in autism. 2022.09.15.508118 Preprint at <https://doi.org/10.1101/2022.09.15.508118> (2022).
123. Paulsen, B. *et al.* Autism genes converge on asynchronous development of shared neuron classes. *Nature* **602**, 268–273 (2022).

124. Wang, P. *et al.* CRISPR/Cas9-mediated heterozygous knockout of the autism gene CHD8 and characterization of its transcriptional networks in cerebral organoids derived from iPS cells. *Mol. Autism* **8**, 11 (2017).
125. Zhang, W. *et al.* Cerebral organoid and mouse models reveal a RAB39b-PI3K-mTOR pathway-dependent dysregulation of cortical development leading to macrocephaly/autism phenotypes. *Genes Dev.* **34**, 580–597 (2020).
126. Sun, X.-Y. *et al.* Generation of vascularized brain organoids to study neurovascular interactions. *eLife* **11**, e76707 (2022).
127. Giandomenico, S. L., Sutcliffe, M. & Lancaster, M. A. Generation and long-term culture of advanced cerebral organoids for studying later stages of neural development. *Nat. Protoc.* **16**, 579–602 (2021).
128. Eisen, J. S. Chapter 1 - History of Zebrafish Research. in *The Zebrafish in Biomedical Research* (eds. Cartner, S. C. *et al.*) 3–14 (Academic Press, 2020). doi:10.1016/B978-0-12-812431-4.00001-4.
129. Parichy, D. M. Advancing biology through a deeper understanding of zebrafish ecology and evolution. *eLife* **4**, e05635 (2015).
130. Dreosti, E., Lopes, G., Kampff, A. & Wilson, S. Development of social behavior in young zebrafish. *Front. Neural Circuits* **9**, (2015).
131. Spence, R., Gerlach, G., Lawrence, C. & Smith, C. The behaviour and ecology of the zebrafish, *Danio rerio*. *Biol. Rev. Camb. Philos. Soc.* **83**, 13–34 (2008).
132. Abril-de-Abreu, R., Cruz, A. S. & Oliveira, R. F. Social dominance modulates eavesdropping in zebrafish. *R. Soc. Open Sci.* **2**, 150220 (2015).
133. Shams, S., Rihel, J., Ortiz, J. G. & Gerlai, R. The zebrafish as a promising tool for modeling human brain disorders: A review based upon an IBNS Symposium. *Neurosci. Biobehav. Rev.* **85**, 176–190 (2018).
134. Zempo, B., Tanaka, N., Daikoku, E. & Ono, F. High-speed camera recordings uncover previously unidentified elements of zebrafish mating behaviors integral to successful fertilization. *Sci. Rep.* **11**, 20228 (2021).
135. Spence, R. & Smith, C. Mating preference of female zebrafish, *Danio rerio*, in relation to male dominance. *Behav. Ecol.* **17**, 779–783 (2006).
136. King, A. C., Gut, M. & Zenker, A. K. Shedding new light on early sex determination in zebrafish. *Arch. Toxicol.* **94**, 4143–4158 (2020).
137. Kobayashi, Y., Nagahama, Y. & Nakamura, M. Diversity and Plasticity of Sex Determination and Differentiation in Fishes. *Sex. Dev.* **7**, 115–125 (2012).
138. Liew, W. C. *et al.* Polygenic Sex Determination System in Zebrafish. *PLoS ONE* **7**, e34397 (2012).
139. Yossa, R. *et al.* A Practical Approach for Sexing Zebrafish, *Danio rerio*. *J. Appl. Aquac.* **25**, (2013).
140. Kossack, M. E. & Draper, B. W. Genetic regulation of sex determination and maintenance in zebrafish (*Danio rerio*). *Curr. Top. Dev. Biol.* **134**, 119–149 (2019).
141. Kishi, S., Slack, B. E., Uchiyama, J. & Zhdanova, I. V. Zebrafish as a Genetic Model in Biological and Behavioral Gerontology: Where Development Meets Aging in Vertebrates – A Mini-Review. *Gerontology* **55**, 430 (2009).
142. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **203**, 253–310 (1995).

143. Parichy, D. M., Elizondo, M. R., Mills, M. G., Gordon, T. N. & Engeszer, R. E. Normal table of postembryonic zebrafish development: staging by externally visible anatomy of the living fish. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **238**, 2975–3015 (2009).
144. Li, J. & Ge, W. Zebrafish as a model for studying ovarian development: Recent advances from targeted gene knockout studies. *Mol. Cell. Endocrinol.* **507**, 110778 (2020).
145. Li, H.-H., Huang, P., Dong, W., Zhu, Z.-Y. & Liu, D. [A brief history of zebrafish research--toward biomedicine]. *Yi Chuan Hered.* **35**, 410–420 (2013).
146. Streisinger, G., Walker, C., Dower, N., Knauber, D. & Singer, F. Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature* **291**, 293–296 (1981).
147. Choi, T.-Y., Choi, T.-I., Lee, Y.-R., Choe, S.-K. & Kim, C.-H. Zebrafish as an animal model for biomedical research. *Exp. Mol. Med.* **53**, 310–317 (2021).
148. Sakai, C., Ijaz, S. & Hoffman, E. J. Zebrafish Models of Neurodevelopmental Disorders: Past, Present, and Future. *Front. Mol. Neurosci.* **11**, 294 (2018).
149. Guo, S. Using zebrafish to assess the impact of drugs on neural development and function. *Expert Opin. Drug Discov.* **4**, 715–726 (2009).
150. Meshalkina, D. A. *et al.* Zebrafish models of autism spectrum disorder. *Exp. Neurol.* **299**, 207–216 (2018).
151. Howe, K. *et al.* The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
152. Stewart, A. M., Nguyen, M., Wong, K., Poudel, M. K. & Kalueff, A. V. Developing zebrafish models of autism spectrum disorder (ASD). *Prog. Neuropsychopharmacol. Biol. Psychiatry* **50**, 27–36 (2014).
153. Rea, V. & Van Raay, T. J. Using Zebrafish to Model Autism Spectrum Disorder: A Comparison of ASD Risk Genes Between Zebrafish and Their Mammalian Counterparts. *Front. Mol. Neurosci.* **13**, 575575 (2020).
154. Canavello, P. R. *et al.* Measuring Endocrine (Cortisol) Responses of Zebrafish to Stress. in *Zebrafish Neurobehavioral Protocols* (eds. Kalueff, A. V. & Cachat, J. M.) 135–142 (Humana Press, 2011). doi:10.1007/978-1-60761-953-6_11.
155. Hisano, Y., Ota, S. & Kawahara, A. Genome editing using artificial site-specific nucleases in zebrafish. *Dev. Growth Differ.* **56**, 26–33 (2014).
156. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229 (2013).
157. Corey, D. R. & Abrams, J. M. Morpholino antisense oligonucleotides: tools for investigating vertebrate development. *Genome Biol.* **2**, reviews1015.1-reviews1015.3 (2001).
158. Stainier, D. Y. R. *et al.* Guidelines for morpholino use in zebrafish. *PLOS Genet.* **13**, e1007000 (2017).
159. Bill, B. R., Petzold, A. M., Clark, K. J., Schimmenti, L. A. & Ekker, S. C. A Primer for Morpholino Use in Zebrafish. *Zebrafish* **6**, 69–77 (2009).
160. Eisen, J. S. & Smith, J. C. Controlling morpholino experiments: don't stop making antisense. *Dev. Camb. Engl.* **135**, 1735–1743 (2008).
161. Kuroyanagi, M. *et al.* New approach for fish breeding by chemical mutagenesis: establishment of TILLING method in fugu (*Takifugu rubripes*) with ENU mutagenesis. *BMC Genomics* **14**, 786 (2013).
162. Moens, C. B., Donn, T. M., Wolf-Saxon, E. R. & Ma, T. P. Reverse genetics in zebrafish by TILLING. *Brief. Funct. Genomic. Proteomic.* **7**, 454–459 (2008).
163. Chang, N. *et al.* Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. *Cell Res.* **23**, 465–472 (2013).

164. Vejnar, C. E., Moreno-Mateos, M. A., Cifuentes, D., Bazzini, A. A. & Giraldez, A. J. Optimized CRISPR-Cas9 System for Genome Editing in Zebrafish. *Cold Spring Harb. Protoc.* **2016**, (2016).
165. Sorlien, E. L., Witucki, M. A. & Ogas, J. Efficient Production and Identification of CRISPR/Cas9-generated Gene Knockouts in the Model System *Danio rerio*. *JoVE J. Vis. Exp.* e56969 (2018) doi:10.3791/56969.
166. Burger, A. *et al.* Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development* **143**, 2025–2037 (2016).
167. Liu, K., Petree, C., Requena, T., Varshney, P. & Varshney, G. K. Expanding the CRISPR Toolbox in Zebrafish for Studying Development and Disease. *Front. Cell Dev. Biol.* **7**, (2019).
168. Kroll, F. *et al.* A simple and effective F0 knockout method for rapid screening of behaviour and other complex phenotypes. *eLife* **10**, e59683 (2021).
169. Masselink, W. Crispants take the spotlight. *Lab Anim.* **50**, 95–96 (2021).
170. Bek, J. W. *et al.* Lrp5 Mutant and Crispant Zebrafish Faithfully Model Human Osteoporosis, Establishing the Zebrafish as a Platform for CRISPR-Based Functional Screening of Osteoporosis Candidate Genes. *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* **36**, 1749–1764 (2021).
171. Moravec, C. E., Voit, G. C., Otterlee, J. & Pelegri, F. Identification of maternal-effect genes in zebrafish using maternal crispants. *Dev. Camb. Engl.* **148**, dev199536 (2021).
172. Diamond, K. M. *et al.* Examining craniofacial variation among crispant and mutant zebrafish models of human skeletal diseases. *J. Anat.* **243**, 66–77 (2023).
173. Buglo, E. *et al.* Genetic compensation in a stable slc25a46 mutant zebrafish: A case for using F0 CRISPR mutagenesis to study phenotypes caused by inherited disease. *PloS One* **15**, e0230566 (2020).
174. Gong, P. *et al.* Novel biallelic loss of EEF1B2 function links to autosomal recessive intellectual disability. *Hum. Mutat.* **43**, 299–304 (2022).
175. Mucci, S. *et al.* Protocol for morphometric analysis of neurons derived from human pluripotent stem cells. *STAR Protoc.* **3**, 101487 (2022).
176. Autar, K. *et al.* A functional hiPSC-cortical neuron differentiation and maturation model and its application to neurological disorders. *Stem Cell Rep.* **17**, 96–109 (2022).
177. Higdon, R. *et al.* The promise of multi-omics and clinical data integration to identify and target personalized healthcare approaches in autism spectrum disorders. *Omics J. Integr. Biol.* **19**, 197–208 (2015).
178. Nomura, J., Mardo, M. & Takumi, T. Molecular signatures from multi-omics of autism spectrum disorders and schizophrenia. *J. Neurochem.* **159**, 647–659 (2021).
179. Antunes, A. S. L. M. & Martins-de-Souza, D. Single-Cell RNA Sequencing and Its Applications in the Study of Psychiatric Disorders. *Biol. Psychiatry Glob. Open Sci.* **3**, 329–339 (2023).
180. Eberwine, J., Sul, J.-Y., Bartfai, T. & Kim, J. The promise of single-cell sequencing. *Nat. Methods* **11**, 25–27 (2014).
181. Wang, P., Zhao, D., Lachman, H. & Zheng, D. Enriched expression of genes associated with autism spectrum disorders in human inhibitory neurons. *Transl. Psychiatry* **8**, (2017).
182. Rubenstein, J. L. R. & Merzenich, M. M. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255–267 (2003).
183. Bozzi, Y., Provenzano, G. & Casarosa, S. Neurobiological bases of autism-epilepsy comorbidity: a focus on excitation/inhibition imbalance. *Eur. J. Neurosci.* **47**, 534–548 (2018).

184. Szoko, N., McShane, A. J. & Natowicz, M. R. Proteomic explorations of autism spectrum disorder. *Autism Res. Off. J. Int. Soc. Autism Res.* **10**, 1460–1469 (2017).
185. Mahony, C. & O’Ryan, C. Convergent Canonical Pathways in Autism Spectrum Disorder from Proteomic, Transcriptomic and DNA Methylation Data. *Int. J. Mol. Sci.* **22**, 10757 (2021).
186. Abraham, J., Szoko, N. & Natowicz, M. R. Proteomic Investigations of Autism Spectrum Disorder: Past Findings, Current Challenges, and Future Prospects. *Adv. Exp. Med. Biol.* **1118**, 235–252 (2019).
187. Murtaza, N., Uy, J. & Singh, K. K. Emerging proteomic approaches to identify the underlying pathophysiology of neurodevelopmental and neurodegenerative disorders. *Mol. Autism* **11**, 27 (2020).
188. Fingleton, E., Li, Y. & Roche, K. W. Advances in Proteomics Allow Insights Into Neuronal Proteomes. *Front. Mol. Neurosci.* **14**, 647451 (2021).
189. Näslund, J. A simple non-invasive method for measuring gross brain size in small live fish with semi-transparent heads. *PeerJ* **2**, e586 (2014).
190. Liu, C.-X. *et al.* CRISPR/Cas9-induced shank3b mutant zebrafish display autism-like behaviors. *Mol. Autism* **9**, 23 (2018).
191. van der Vaart, M. *et al.* Mecp2 regulates tnfa during zebrafish embryonic development and acute inflammation. *Dis. Model. Mech.* **10**, 1439–1451 (2017).
192. Patten, S. A. *et al.* Role of Chd7 in Zebrafish: A Model for CHARGE Syndrome. *PLoS ONE* **7**, e31650 (2012).
193. Bernier, R. *et al.* Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* **158**, 263–276 (2014).
194. Tayanloo-Beik, A. *et al.* Zebrafish Modeling of Autism Spectrum Disorders, Current Status and Future Prospective. *Front. Psychiatry* **13**, 911770 (2022).
195. Norton, W. H. J., Manceau, L. & Reichmann, F. The Visually Mediated Social Preference Test: A Novel Technique to Measure Social Behavior and Behavioral Disturbances in Zebrafish. *Methods Mol. Biol. Clifton NJ* **2011**, 121–132 (2019).
196. Ariyasiri, K. *et al.* Pharmacological (ethanol) and mutation (sam2 KO) induced impairment of novelty preference in zebrafish quantified using a new three-chamber social choice task. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **88**, 53–65 (2019).
197. Zabegalov, K. N. *et al.* Abnormal repetitive behaviors in zebrafish and their relevance to human brain disorders. *Behav. Brain Res.* **367**, 101–110 (2019).
198. Schnörr, S. J., Steenbergen, P. J., Richardson, M. K. & Champagne, D. L. Measuring thigmotaxis in larval zebrafish. *Behav. Brain Res.* **228**, 367–374 (2012).
199. Grone, B. P., Qu, T. & Baraban, S. C. Behavioral Comorbidities and Drug Treatments in a Zebrafish scn1lab Model of Dravet Syndrome. *eNeuro* **4**, ENEURO.0066-17.2017 (2017).
200. Pietri, T. *et al.* The first mecp2-null zebrafish model shows altered motor behaviors. *Front. Neural Circuits* **7**, 118 (2013).
201. Hoffman, E. J. *et al.* Estrogens Suppress a Behavioral Phenotype in Zebrafish Mutants of the Autism Risk Gene, CNTNAP2. *Neuron* **89**, 725–733 (2016).
202. Hong, D. & Iakoucheva, L. M. Therapeutic strategies for autism: targeting three levels of the central dogma of molecular biology. *Transl. Psychiatry* **13**, 1–14 (2023).
203. Ghosh, A., Nadella, N., Monaghan-Nichols, A. P. & Chu, X.-P. Gene therapy as an emerging treatment for Scn2a mutation-induced autism spectrum disorders. *Fundam. Res.* (2023) doi:10.1016/j.fmre.2023.02.004.

204. European Medicines Agency (EMA). Guideline on the clinical development of medicinal products for the treatment of Autism Spectrum Disorder (ASD). (2017).
205. Weuring, W., Geerligs, J. & Koeleman, B. P. C. Gene Therapies for Monogenic Autism Spectrum Disorders. *Genes* **12**, 1667 (2021).
206. Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995 (2022).
207. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
208. Gao, Y. *et al.* Concurrent mutations associated with trastuzumab-resistance revealed by single cell sequencing. *Breast Cancer Res. Treat.* **187**, 613–624 (2021).
209. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
210. Xu, J. *et al.* Molecular cloning and characterization of a novel human BTBD8 gene containing double BTB/POZ domains. *Int. J. Mol. Med.* **13**, 193–197 (2004).
211. Piccini, A. *et al.* APACHE Is an AP2-Interacting Protein Involved in Synaptic Vesicle Trafficking and Neuronal Development. *Cell Rep.* **21**, 3596–3611 (2017).
212. Sigrist, C. J. A. *et al.* New and continuing developments at PROSITE. *Nucleic Acids Res.* **41**, D344–D347 (2013).
213. Bonchuk, A., Balagurov, K. & Georgiev, P. BTB domains: A structural view of evolution, multimerization, and protein–protein interactions. *BioEssays* **45**, 2200179 (2023).
214. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
215. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
216. Alonso González, A. Bases genéticas de los trastornos del espectro autista: estudio de la variación común y rara. (2020).
217. Pankratz, N. *et al.* Meta-analysis of rare and common exome chip variants identifies S1PR4 and other loci influencing blood cell traits. *Nat. Genet.* **48**, 867–876 (2016).
218. Huo, X. *et al.* Prognostic factors and Doxorubicin involved in malignant progression of meningioma. *Sci. Rep.* **13**, 5632 (2023).
219. Seal, R. L. *et al.* Genenames.org: the HGNC resources in 2023. *Nucleic Acids Res.* **51**, D1003–D1009 (2023).
220. Molven, A., Houge, G. & Berger, R. Chromosomal Assignment of the Human Gene Encoding the Fos-Related Antigen-2 (FRA2) to Chromosome 2p22–p23. *Genomics* **38**, 72–75 (1996).
221. Cospain, A. *et al.* FOSL2 truncating variants in the last exon cause a neurodevelopmental disorder with scalp and enamel defects. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **24**, 2475–2486 (2022).
222. Bejjani, F., Evanno, E., Zibara, K., Piechaczyk, M. & Jariel-Encontre, I. The AP-1 transcriptional complex: Local switch or remote command? *Biochim. Biophys. Acta BBA - Rev. Cancer* **1872**, 11–23 (2019).
223. Li, S., Fang, X.-D., Wang, X.-Y. & Fei, B.-Y. Fos-like antigen 2 (FOSL2) promotes metastasis in colon cancer. *Exp. Cell Res.* **373**, 57–61 (2018).
224. Wan, X. *et al.* FOSL2 promotes VEGF-independent angiogenesis by transcriptionally activating Wnt5a in breast cancer-associated fibroblasts. *Theranostics* **11**, 4975–4991 (2021).
225. Sarode, P. *et al.* Reprogramming of tumor-associated macrophages by targeting β -catenin/FOSL2/ARID5A signaling: A potential treatment of lung cancer. *Sci. Adv.* **6**, eaaz6105 (2020).

226. Wu, L. *et al.* Natural Coevolution of Tumor and Immunoenvironment in Glioblastoma. *Cancer Discov.* **12**, 2820–2837 (2022).
227. Bian, Z. *et al.* SNHG17 promotes colorectal tumorigenesis and metastasis via regulating Trim23-PES1 axis and miR-339-5p-FOSL2-SNHG17 positive feedback loop. *J. Exp. Clin. Cancer Res. CR* **40**, 360 (2021).
228. Wang, J. *et al.* FOSL2 positively regulates TGF- β 1 signalling in non-small cell lung cancer. *PLoS One* **9**, e112150 (2014).
229. Li, J., Zhou, L., Jiang, H., Lin, L. & Li, Y. Inhibition of FOSL2 aggravates the apoptosis of ovarian cancer cells by promoting the formation of inflammasomes. *Genes Genomics* **44**, 29–38 (2022).
230. Li, S., Liu, Z., Fang, X.-D., Wang, X.-Y. & Fei, B.-Y. MicroRNA (miR)-597-5p Inhibits Colon Cancer Cell Migration and Invasion by Targeting FOS-Like Antigen 2 (FOSL2). *Front. Oncol.* **9**, 495 (2019).
231. He, J. *et al.* miR-597 inhibits breast cancer cell proliferation, migration and invasion through FOSL2. *Oncol. Rep.* **37**, 2672–2678 (2017).
232. Shetty, A. *et al.* Interactome Networks of FOSL1 and FOSL2 in Human Th17 Cells. *ACS Omega* **6**, 24834–24847 (2021).
233. Ubieto, K. *et al.* Fra-2 regulates B cell development by enhancing IRF4 and Foxo1 transcription. *J. Exp. Med.* **214**, 2059–2071 (2017).
234. Cabrera-Ortega, A. A., Feinberg, D., Liang, Y., Rossa, C. & Graves, D. T. The Role of Forkhead Box 1 (FOXO1) in the Immune System: Dendritic Cells, T Cells, B Cells, and Hematopoietic Stem Cells. *Crit. Rev. Immunol.* **37**, 1–13 (2017).
235. Renoux, F. *et al.* The AP1 Transcription Factor FosL2 Promotes Systemic Autoimmunity and Inflammation by Repressing Treg Development. *Cell Rep.* **31**, 107826 (2020).
236. Bozec, A. *et al.* Osteoclast size is controlled by Fra-2 through LIF/LIF-receptor signalling and hypoxia. *Nature* **454**, 221–225 (2008).
237. Nishikawa, S. Localization of transcription factor AP-1 family proteins in ameloblast nuclei of the rat incisor. *J. Histochem. Cytochem. Off. J. Histochem. Soc.* **48**, 1511–1520 (2000).
238. Jahangiri, L. *et al.* The AP-1 transcription factor component Fosl2 potentiates the rate of myocardial differentiation from the zebrafish second heart field. *Dev. Camb. Engl.* **143**, 113–122 (2016).
239. Seidenberg, J. *et al.* The AP-1 Transcription Factor Fosl-2 Regulates Autophagy in Cardiac Fibroblasts during Myocardial Fibrogenesis. *Int. J. Mol. Sci.* **22**, 1861 (2021).
240. Ucerro, A. C. *et al.* Fra-2-expressing macrophages promote lung fibrosis in mice. *J. Clin. Invest.* **129**, 3293–3309 (2019).
241. Gao, F. *et al.* Heterozygous mutations in SMARCA2 reprogram the enhancer landscape by global retargeting of SMARCA4. *Mol. Cell* **75**, 891-904.e7 (2019).
242. Wenderski, W. *et al.* Loss of the neural-specific BAF subunit ACTL6B relieves repression of early response genes and causes recessive autism. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 10055–10066 (2020).
243. White, R. J. *et al.* A high-resolution mRNA expression time course of embryonic development in zebrafish. *eLife* **6**, e30860 (2017).
244. Raman, M., Earnest, S., Zhang, K., Zhao, Y. & Cobb, M. H. TAO kinases mediate activation of p38 in response to DNA damage. *EMBO J.* **26**, 2005–2014 (2007).
245. Fang, C.-Y., Lai, T.-C., Hsiao, M. & Chang, Y.-C. The Diverse Roles of TAO Kinases in Health and Diseases. *Int. J. Mol. Sci.* **21**, 7463 (2020).

246. Hunter, J. M. *et al.* Inherited and de novo variants extend the etiology of TAOK1-associated neurodevelopmental disorder. *Mol. Case Stud.* mcs.a006180 (2022) doi:10.1101/mcs.a006180.
247. Ye, J., Shi, M., Chen, W., Zhu, F. & Duan, Q. Research Advances in the Molecular Functions and Relevant Diseases of TAOs, Novel STE20 Kinase Family Members. *Curr. Pharm. Des.* **26**, 3122–3133 (2020).
248. Woerden, G. M. *et al.* TAOK1 is associated with neurodevelopmental disorder and essential for neuronal maturation and cortical development. *Hum. Mutat.* **42**, 445–459 (2021).
249. Plouffe, S. W. *et al.* Characterization of Hippo Pathway Components by Gene Inactivation. *Mol. Cell* **64**, 993–1008 (2016).
250. Koo, C.-Y. *et al.* Targeting TAO Kinases Using a New Inhibitor Compound Delays Mitosis and Induces Mitotic Cell Death in Centrosome Amplified Breast Cancer Cells. *Mol. Cancer Ther.* **16**, 2410–2421 (2017).
251. Liskovych, M. *et al.* A novel assay to screen siRNA libraries identifies protein kinases required for chromosome transmission. *Genome Res.* **29**, 1719–1732 (2019).
252. Timm, T. MARKK, a Ste20-like kinase, activates the polarity-inducing kinase MARK/PAK-1. *EMBO J.* **22**, 5090–5101 (2003).
253. Dulovic-Mahlow, M. *et al.* De Novo Variants in TAOK1 Cause Neurodevelopmental Disorders. *Am. J. Hum. Genet.* **105**, 213–220 (2019).
254. King, I. & Heberlein, U. Tao kinases as coordinators of actin and microtubule dynamics in developing neurons. *Commun. Integr. Biol.* **4**, 554–556 (2011).
255. Ultanir, S. K. *et al.* MST3 kinase phosphorylates TAO1/2 to enable Myosin Va function in promoting spine synapse development. *Neuron* **84**, 968–982 (2014).
256. Beeman, N., Sapre, T., Ong, S.-E. & Yadav, S. Neurodevelopmental disorder-associated mutations in TAOK1 reveal its function as a plasma membrane remodeling kinase. *Sci. Signal.* **16**, eadd3269 (2023).
257. Zhang, Z. *et al.* TAOK1 negatively regulates IL-17-mediated signaling and inflammation. *Cell. Mol. Immunol.* **15**, 794–802 (2018).
258. Zhu, L. *et al.* TAOK1 positively regulates TLR4-induced inflammatory responses by promoting ERK1/2 activation in macrophages. *Mol. Immunol.* **122**, 124–131 (2020).
259. Luo, X. *et al.* Ste20-Like Kinase TAOK1 Positively Regulates Antiviral Responses by Controlling the TBK1-IRF3 Signaling Axis. *J. Innate Immun.* **15**, 1–17 (2023).
260. Pennemann, F. L. *et al.* Cross-species analysis of viral nucleic acid interacting proteins identifies TAOs as innate immune regulators. *Nat. Commun.* **12**, 7009 (2021).
261. Zihni, C., Mitsopoulos, C., Tavares, I. A., Ridley, A. J. & Morris, J. D. H. Prostate-derived Sterile 20-like Kinase 2 (PSK2) Regulates Apoptotic Morphology via C-Jun N-terminal Kinase and Rho Kinase-1. *J. Biol. Chem.* **281**, 7317–7323 (2006).
262. Chen, L. TAOK1 Promotes Proliferation and Invasion of Non-Small-Cell Lung Cancer Cells by Inhibition of WWC1. *Comput. Math. Methods Med.* **2022**, 3157448 (2022).
263. Rattanapan, Y. *et al.* High Expression of miR-483-5p Predicts Chemotherapy Resistance in Epithelial Ovarian Cancer. *MicroRNA Shariqah United Arab Emir.* **10**, 51–57 (2021).
264. Schulte, I. *et al.* Structural analysis of the genome of breast cancer cell line ZR-75-30 identifies twelve expressed fusion genes. *BMC Genomics* **13**, 719 (2012).
265. Cai, C., Zeng, Q., Zhou, G. & Mu, X. Identification of novel transcription factor-microRNA-mRNA co-regulatory networks in pulmonary large-cell neuroendocrine carcinoma. *Ann. Transl. Med.* **9**, 133 (2021).

266. Beg, A., Parveen, R., Fouad, H., Yahia, M. E. & Hassanein, A. S. Identification of Driver Genes and miRNAs in Ovarian Cancer through an Integrated In-Silico Approach. *Biology* **12**, 192 (2023).
267. Vélez-Reyes, G. L. *et al.* Transposon Mutagenesis-Guided CRISPR/Cas9 Screening Strongly Implicates Dysregulation of Hippo/YAP Signaling in Malignant Peripheral Nerve Sheath Tumor Development. *Cancers* **13**, 1584 (2021).
268. Giacomini, C. *et al.* A new TAO kinase inhibitor reduces tau phosphorylation at sites associated with neurodegeneration in human tauopathies. *Acta Neuropathol. Commun.* **6**, 37 (2018).
269. Zach, S., Felk, S. & Gillardon, F. Signal transduction protein array analysis links LRRK2 to Ste20 kinases and PKC zeta that modulate neuronal plasticity. *PloS One* **5**, e13191 (2010).
270. Xie, B. *et al.* A novel de novo microdeletion at 17q11.2 adjacent to NF1 gene associated with developmental delay, short stature, microcephaly and dysmorphic features. *Mol. Cytogenet.* **9**, 41 (2016).
271. Yu, L. *et al.* Paternal De Novo Variant of TAOK1 in a Fetus With Structural Brain Abnormalities. *Front. Genet.* **13**, 836853 (2022).
272. Sampedro-Castañeda, M. & Ultanir, S. K. Activity-dependent membrane sculpting deficits in TAOK1-linked neurodevelopmental disease. *Trends Neurosci.* **46**, 611–613 (2023).
273. Blaschek, A. *et al.* TAOK1-related neurodevelopmental disorder: A new differential diagnosis for childhood-onset tremor? *Parkinsonism Relat. Disord.* **109**, 105320 (2023).
274. Ministerio de la Presidencia. *Real Decreto 53/2013, de 1 de febrero, por el que se establecen las normas básicas aplicables para la protección de los animales utilizados en experimentación y otros fines científicos, incluyendo la docencia.* vol. BOE-A-2013-1337 11370–11421 (2013).
275. Ministerio de la Presidencia, Relaciones con las Cortes e Igualdad. *Real Decreto 1386/2018, de 19 de noviembre, por el que se modifica el Real Decreto 53/2013, de 1 de febrero, por el que se establecen las normas básicas aplicables para la protección de los animales utilizados en experimentación y otros fines científicos, incluyendo la docencia.* vol. BOE-A-2018-15797 112804–112806 (2018).
276. Ministerio de Economía y Competitividad. *Orden ECC/566/2015, de 20 de marzo, por la que se establecen los requisitos de capacitación que debe cumplir el personal que maneje animales utilizados, criados o suministrados con fines de experimentación y otros fines científicos, incluyendo la docencia.* vol. BOE-A-2015-3564 27940–27973 (2015).
277. European Parliament, Council of the European Union. *Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes.* vol. 2008/0211/COD 33–79 (2010).
278. Instituto de Salud Carlos III (ISCIII) & Banco Nacional de Líneas Celulares. Ctrl. BS PBiPS37-Sv4F-1 Cell Line. <https://www.isciii.es/QueHacemos/Servicios/BIOBANCOS/BNLC/Paginas/LineasiPS.aspx> (2022).
279. Canals, I. & Ahlenius, H. CRISPR/Cas9 Genome Engineering in Human Pluripotent Stem Cells for Modeling of Neurological Disorders. in *Neural Reprogramming* (ed. Ahlenius, H.) vol. 2352 237–251 (Springer US, 2021).
280. Lei Stanley Qi Lab & Xiaowo Wang Lab. CRISPR-ERA Version 1.2. *CRISPR-ERA: a comprehensive designer tool for CRISPR genome editing, (gene) repression, and activation* <http://crispr-era.stanford.edu/> (2014).

281. Synthego. Synthego's Knockout Guide Design v1.3. <https://design.synthego.com/#/> (2023).
282. Synthego. Synthego's Verify Guide Design webtool v1.3. <https://design.synthego.com/#/validate> (2023).
283. Delrue, I., Pan, Q., Baczmanska, A. K., Callens, B. W. & Verdoodt, L. L. M. Determination of the Selection Capacity of Antibiotics for Gene Selection. *Biotechnol. J.* **13**, e1700747 (2018).
284. Synthego. Synthego's ICE analysis v3.0. <https://ice.synthego.com/#/> (2023).
285. Instituto de Salud Carlos III (ISCIII) & Banco Nacional de Líneas Celulares. Nomenclatura para las líneas hiPSCs y líneas genéticamente modificadas. *Web Instituto de Salud Carlos III*. <https://www.isciii.es/QueHacemos/Servicios/BIOBANCOS/BNLC/Documents/2021%20-0-NOMENCLATURA%20PARA%20LA%20L%C3%8DNEAS%20IPS%20y%20G%20modificadas.pdf>.
286. Velasco, S., Paulsen, B. & Arlotta, P. Highly reproducible human brain organoids recapitulate cerebral cortex cellular diversity. (2019) doi:10.21203/rs.2.9542/v1.
287. Campeau, E. *et al.* A versatile viral system for expression and depletion of proteins in mammalian cells. *PloS One* **4**, e6529 (2009).
288. Meijering, E. *et al.* Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytom. Part J. Int. Soc. Anal. Cytol.* **58**, 167–176 (2004).
289. Fonseka, P., Pathan, M., Chitti, S. V., Kang, T. & Mathivanan, S. FunRich enables enrichment analysis of OMICS datasets. *J. Mol. Biol.* **433**, 166747 (2021).
290. STRING: functional protein association networks. https://string-db.org/cgi/input?sessionId=bjj6ZDWCPLWh&input_page_show_search=on.
291. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
292. Tanaka, Y., Cakir, B., Xiang, Y., Sullivan, G. J. & Park, I.-H. Synthetic Analyses of Single-Cell Transcriptomes from Multiple Brain Organoids and Fetal Brain. *Cell Rep.* **30**, 1682-1689.e3 (2020).
293. Cowan, C. S. *et al.* Cell Types of the Human Retina and Its Organoids at Single-Cell Resolution. *Cell* **182**, 1623-1640.e34 (2020).
294. Velmeshev, D. *et al.* Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019).
295. Bhaduri, A. *et al.* Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142–148 (2020).
296. Speir, M. L. *et al.* UCSC Cell Browser: visualize your single-cell data. *Bioinformatics* **37**, 4578–4580 (2021).
297. Uzquiano, A. *et al.* Proper acquisition of cell class identity in organoids allows definition of fate specification programs of the human cerebral cortex. *Cell* **185**, 3770-3788.e27 (2022).
298. Moreno-Mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR/Cas9 targeting in vivo. *Nat. Methods* **12**, 982–988 (2015).
299. Jiang, S. *et al.* Integrative system biology analyses of CRISPR-edited iPSC-derived neurons and human brains reveal deficiencies of presynaptic signaling in FTL and PSP. *Transl. Psychiatry* **8**, 1–16 (2018).
300. Kubra, K. *et al.* Phylogenetic and Expression Analysis of Fos Transcription Factors in Zebrafish. *Int. J. Mol. Sci.* **23**, 10098 (2022).

301. Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).
302. Lipovich, L. *et al.* High-throughput RNA sequencing reveals structural differences of orthologous brain-expressed genes between western lowland gorillas and humans. *J. Comp. Neurol.* **524**, 288–308 (2016).
303. Chan, W. K., Griffiths, R., Price, D. J. & Mason, J. O. Cerebral organoids as tools to identify the developmental roots of autism. *Mol. Autism* **11**, 58 (2020).
304. Kilpatrick, S., Irwin, C. & Singh, K. K. Human pluripotent stem cell (hPSC) and organoid models of autism: opportunities and limitations. *Transl. Psychiatry* **13**, 1–21 (2023).
305. Hazlett, H. C. *et al.* Early brain development in infants at high risk for autism spectrum disorder. *Nature* **542**, 348–351 (2017).
306. Sacco, R., Gabriele, S. & Persico, A. M. Head circumference and brain size in autism spectrum disorder: A systematic review and meta-analysis. *Psychiatry Res.* **234**, 239–251 (2015).
307. Courchesne, E., Carper, R. & Akshoomoff, N. Evidence of brain overgrowth in the first year of life in autism. *JAMA* **290**, 337–344 (2003).
308. Courchesne, E. *et al.* Neuron number and size in prefrontal cortex of children with autism. *JAMA* **306**, 2001–2010 (2011).
309. Florio, M. & Huttner, W. B. Neural progenitors, neurogenesis and the evolution of the neocortex. *Dev. Camb. Engl.* **141**, 2182–2194 (2014).
310. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
311. Butler, M. G. *et al.* Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *J. Med. Genet.* **42**, 318–321 (2005).
312. Patak, J., Zhang-James, Y. & Faraone, S. V. Endosomal system genetics and autism spectrum disorders: A literature review. *Neurosci. Biobehav. Rev.* **65**, 95–112 (2016).
313. Ohno, H. Physiological Roles of Clathrin Adaptor AP Complexes: Lessons from Mutant Animals. *J. Biochem. (Tokyo)* **139**, 943–948 (2006).
314. Litwa, K., Seong, E., Burmeister, M. & Faundez, V. Neuronal and non-neuronal functions of the AP-3 sorting machinery. *J. Cell Sci.* **120**, 531–41 (2007).
315. Grabner, C. P., Price, S. D., Lysakowski, A., Cahill, A. L. & Fox, A. P. Regulation of large dense-core vesicle volume and neurotransmitter content mediated by adaptor protein 3. *Proc. Natl. Acad. Sci.* **103**, 10035–10040 (2006).
316. Shetty, A. *et al.* The neural cell adhesion molecule promotes maturation of the presynaptic endocytotic machinery by switching synaptic vesicle recycling from adaptor protein 3 (AP-3)- to AP-2-dependent mechanisms. *J. Neurosci. Off. J. Soc. Neurosci.* **33**, 16828–16845 (2013).
317. Hao, W., Luo, Z., Zheng, L., Prasad, K. & Lafer, E. M. AP180 and AP-2 interact directly in a complex that cooperatively assembles clathrin. *J. Biol. Chem.* **274**, 22785–22794 (1999).
318. Petralia, R. S. *et al.* Reduction of AP180 and CALM produces defects in synaptic vesicle size and density. *Neuromolecular Med.* **15**, 49–60 (2013).
319. Robinson, M. S. Adaptable adaptors for coated vesicles. *Trends Cell Biol.* **14**, 167–174 (2004).
320. Tariq, K. & Luikart, B. W. Striking a balance: PIP2 and PIP3 signaling in neuronal health and disease. *Explor. Neuroprotective Ther.* **1**, 86–100 (2021).

321. Durieux, A.-C. *et al.* A centronuclear myopathy-dynamin 2 mutation impairs skeletal muscle structure and function in mice. *Hum. Mol. Genet.* **19**, 4820–4836 (2010).
322. Bitoun, M. *et al.* Dynamin 2 mutations associated with human diseases impair clathrin-mediated receptor endocytosis. *Hum. Mutat.* **30**, 1419–1427 (2009).
323. Deneubourg, C. *et al.* The spectrum of neurodevelopmental, neuromuscular and neurodegenerative disorders due to defective autophagy. *Autophagy* **18**, 496–517 (2022).
324. Zhao, M., Maani, N. & Dowling, J. J. Dynamin 2 (DNM2) as Cause of, and Modifier for, Human Neuromuscular Disease. *Neurotherapeutics* **15**, 966–975 (2018).
325. Gómez-Oca, R. *et al.* Differential impact of ubiquitous and muscle dynamin 2 isoforms in muscle physiology and centronuclear myopathy. *Nat. Commun.* **13**, 6849 (2022).
326. Dana, H., Tahtsakal, R. & Sener, E. F. Animal models of autism: a perspective from autophagy mechanisms. *J. Transl. Genet. Genomics* **4**, 251–262 (2020).
327. Ferguson, S. *et al.* Coordinated Actions of Actin and BAR Proteins Upstream of Dynamin at Endocytic Clathrin-Coated Pits. *Dev. Cell* **17**, 811–822 (2009).
328. Yamashita, T., Hige, T. & Takahashi, T. Vesicle Endocytosis Requires Dynamin-Dependent GTP Hydrolysis at a Fast CNS Synapse. *Science* **307**, 124–127 (2005).
329. Nykänen, N.-P., Kysenius, K., Sakha, P., Tammela, P. & Huttunen, H. J. γ -Aminobutyric Acid Type A (GABAA) Receptor Activation Modulates Tau Phosphorylation. *J. Biol. Chem.* **287**, 6743–6752 (2012).
330. Li, Y. *et al.* Implications of GABAergic Neurotransmission in Alzheimer's Disease. *Front. Aging Neurosci.* **8**, (2016).
331. Chang, C.-W., Evans, M. D., Yu, X., Yu, G.-Q. & Mucke, L. Tau reduction affects excitatory and inhibitory neurons differently, reduces excitation/inhibition ratios, and counteracts network hypersynchrony. *Cell Rep.* **37**, 109855 (2021).
332. Tai, C. *et al.* Tau Reduction Prevents Key Features of Autism in Mouse Models. *Neuron* **106**, 421–437.e11 (2020).
333. Lee, W. S. *et al.* Syntaxins 6 and 8 facilitate tau into secretory pathways. *Biochem. J.* **478**, 1471–1484 (2021).
334. Mori, Y., Takenaka, K.-I., Fukazawa, Y. & Takamori, S. The endosomal Q-SNARE, Syntaxin 7, defines a rapidly replenishing synaptic vesicle recycling pool in hippocampal neurons. *Commun. Biol.* **4**, 981 (2021).
335. Wu, J. *et al.* Syntaxin 7 modulates seizure activity in epilepsy. *Neurobiol. Dis.* **181**, 106118 (2023).
336. Cali, E., Rocca, C., Salpietro, V. & Houlden, H. Epileptic Phenotypes Associated With SNAREs and Related Synaptic Vesicle Exocytosis Machinery. *Front. Neurol.* **12**, 806506 (2021).
337. Lanoue, V. *et al.* STXBP1 encephalopathy: Connecting neurodevelopmental disorders with α -synucleinopathies? *Neurology* **93**, 114–123 (2019).
338. Lopes, F. *et al.* Identification of novel genetic causes of Rett syndrome-like phenotypes. *J. Med. Genet.* **53**, 190–199 (2016).
339. Miyamoto, H. *et al.* Potentiation of excitatory synaptic transmission ameliorates aggression in mice with Stxbp1 haploinsufficiency. *Hum. Mol. Genet.* **26**, 4961–4974 (2017).
340. Wang, C. T. *et al.* Synaptotagmin modulation of fusion pore kinetics in regulated exocytosis of dense-core vesicles. *Science* **294**, 1111–1115 (2001).
341. Tan, J. Z. A. *et al.* Copine-6 is a Ca²⁺ sensor for activity-induced AMPA receptor exocytosis. *Cell Rep.* **42**, 113460 (2023).
342. Schroer, T. A. Dynactin. *Annu. Rev. Cell Dev. Biol.* **20**, 759–779 (2004).

343. Boumil, R. M. *et al.* A Missense Mutation in a Highly Conserved Alternate Exon of Dynamin-1 Causes Epilepsy in Fitful Mice. *PLOS Genet.* **6**, e1001046 (2010).
344. Horiuchi, K. *et al.* Identification of Wilms' Tumor 1-associating Protein Complex and Its Role in Alternative Splicing and the Cell Cycle. *J. Biol. Chem.* **288**, 33292–33302 (2013).
345. Wei, H., Alberts, I. & Li, X. The apoptotic perspective of autism. *Int. J. Dev. Neurosci. Off. J. Int. Soc. Dev. Neurosci.* **36**, 13–18 (2014).
346. Araghi-Niknam, M. & Fatemi, S. H. Levels of Bcl-2 and P53 Are Altered in Superior Frontal and Cerebellar Cortices of Autistic Subjects. *Cell. Mol. Neurobiol.* **23**, 945–952 (2003).
347. Sheikh, A. M. *et al.* BDNF-Akt-Bcl2 antiapoptotic signaling pathway is compromised in the brain of autistic subjects. *J. Neurosci. Res.* **88**, 2641–2647 (2010).
348. Pardo, C. A. & Eberhart, C. G. The Neurobiology of Autism. *Brain Pathol.* **17**, 434–447 (2007).
349. Davis, T. A., Loos, B. & Engelbrecht, A.-M. AHNAK: The giant jack of all trades. *Cell. Signal.* **26**, 2683–2693 (2014).
350. Sundararaj, S., Ravindran, A. & Casarotto, M. G. AHNAK: The quiet giant in calcium homeostasis. *Cell Calcium* **96**, 102403 (2021).
351. Lu, F., Zhu, L., Jia, X., Wang, J. & Mu, P. Downregulated in renal carcinoma 1 (DRR1) mediates the differentiation of neural stem cells through transcriptional regulation. *Neurosci. Lett.* **756**, 135943 (2021).
352. Guerrini, R. *et al.* Phenotypic and genetic spectrum of ATP6V1A encephalopathy: a disorder of lysosomal homeostasis. *Brain J. Neurol.* **145**, 2687–2703 (2022).
353. Werner, A. *et al.* Cell-fate determination by ubiquitin-dependent regulation of translation. *Nature* **525**, 523–527 (2015).
354. Danelon, V. *et al.* Modular and Distinct Plexin-A4/FARP2/Rac1 Signaling Controls Dendrite Morphogenesis. *J. Neurosci. Off. J. Soc. Neurosci.* **40**, 5413–5430 (2020).
355. Eisenberg, C. *et al.* Reduced hippocampal inhibition and enhanced autism-epilepsy comorbidity in mice lacking neuropilin 2. *Transl. Psychiatry* **11**, 1–12 (2021).
356. Khmour, H. Y. *et al.* Neuropilin 2/Plexin-A3 Receptors Regulate the Functional Connectivity and the Excitability in the Layers 4 and 5 of the Cerebral Cortex. *J. Neurosci.* **42**, 4828–4840 (2022).
357. Bal, M. *et al.* Reelin mobilizes a VAMP7-dependent synaptic vesicle pool and selectively augments spontaneous neurotransmission. *Neuron* **80**, 10.1016/j.neuron.2013.08.024 (2013).
358. Elabd, S. *et al.* Delay in development and behavioural abnormalities in the absence of p53 in zebrafish. *PLoS ONE* **14**, e0220069 (2019).
359. Eferl, R., Zenz, R., Theussl, H.-C. & Wagner, E. F. Simultaneous generation of fra-2 conditional and fra-2 knock-out mice. *genesis* **45**, 447–451 (2007).
360. Karreth, F., Hoebertz, A., Scheuch, H., Eferl, R. & Wagner, E. F. The AP1 transcription factor Fra2 is required for efficient cartilage development. *Development* **131**, 5717–5725 (2004).
361. Ciofani, M. *et al.* A validated regulatory network for Th17 cell specification. *Cell* **151**, 289–303 (2012).
362. Asano, Y. Recent advances in animal models of systemic sclerosis. *J. Dermatol.* **43**, 19–28 (2016).
363. Bozec, A. *et al.* Fra-2/AP-1 controls bone formation by regulating osteoblast differentiation and collagen production. *J. Cell Biol.* **190**, 1093–1106 (2010).

364. Huang, L.-C. *et al.* KinOrtho: a method for mapping human kinase orthologs across the tree of life and illuminating understudied kinases. *BMC Bioinformatics* **22**, 446 (2021).
365. Poon, C. L. C., Mitchell, K. A., Kondo, S., Cheng, L. Y. & Harvey, K. F. The Hippo Pathway Regulates Neuroblasts and Brain Size in *Drosophila melanogaster*. *Curr. Biol.* **26**, 1034–1042 (2016).
366. Wang, J. *et al.* Taok1 haploinsufficiency leads to autistic-like behaviors in mice via the dorsal raphe nucleus. *Cell Rep.* **42**, 113078 (2023).
367. Liu, C.-Y., Yang, Y.-H., Li, P.-S., Gan, M.-L. & Li, D.-Z. Fetal macrocephaly in the third trimester: Prenatal phenotype of TAOK1-associated neurodevelopmental disorder. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **289**, 219–220 (2023).
368. Kilb, W. & Frotscher, M. Cajal-Retzius cells: organizers of cortical development. *E-Neuroforum* **7**, 82–88 (2016).
369. Trutzer, I. M., García-Cabezas, M. Á. & Zikopoulos, B. Postnatal development and maturation of layer 1 in the lateral prefrontal cortex and its disruption in autism. *Acta Neuropathol. Commun.* **7**, 40 (2019).
370. Camacho, J., Ejaz, E., Ariza, J., Noctor, S. C. & Martínez-Cerdeño, V. RELN-expressing neuron density in layer I of the superior temporal lobe is similar in human brains with autism and in age-matched controls. *Neurosci. Lett.* **579**, 163–167 (2014).
371. Vílchez-Acosta, A. *et al.* Specific contribution of Reelin expressed by Cajal–Retzius cells or GABAergic interneurons to cortical lamination. *Proc. Natl. Acad. Sci.* **119**, e2120079119 (2022).
372. Fatemi, S. H., Sary, J. M., Halt, A. R. & Realmuto, G. R. Dysregulation of Reelin and Bcl-2 proteins in autistic cerebellum. *J. Autism Dev. Disord.* **31**, 529–535 (2001).
373. Loohuis, N. F. O. *et al.* Altered Expression of Circadian Rhythm and Extracellular Matrix Genes in the Medial Prefrontal Cortex of a Valproic Acid Rat Model of Autism. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **77**, 128–132 (2017).
374. Luo, R. *et al.* G protein-coupled receptor 56 and collagen III, a receptor-ligand pair, regulates cortical development and lamination. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12925–12930 (2011).
375. Hardwick, L. J. A. & Philpott, A. Multi-site phosphorylation regulates NeuroD4 activity during primary neurogenesis: a conserved mechanism amongst proneural proteins. *Neural Develop.* **10**, 15 (2015).
376. Sood, S. *et al.* CHD8 dosage regulates transcription in pluripotency and early murine neural differentiation. *Proc. Natl. Acad. Sci.* **117**, 22331–22340 (2020).
377. Spiga, F. M., Prouteau, M. & Gotta, M. The TAO kinase KIN-18 regulates contractility and establishment of polarity in the *C. elegans* embryo. *Dev. Biol.* **373**, 26–38 (2013).
378. Gzielo, K. & Nikiforuk, A. Astroglia in Autism Spectrum Disorder. *Int. J. Mol. Sci.* **22**, 11544 (2021).
379. Vakilzadeh, G. & Martinez-Cerdeño, V. Pathology and Astrocytes in Autism. *Neuropsychiatr. Dis. Treat.* **19**, 841–850 (2023).
380. Hu, C., Feng, P., Yang, Q. & Xiao, L. Clinical and Neurobiological Aspects of TAO Kinase Family in Neurodevelopmental Disorders. *Front. Mol. Neurosci.* **14**, (2021).
381. Turchiano, G. *et al.* Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. *Cell Stem Cell* **28**, 1136–1147.e5 (2021).
382. Weissbein, U., Benvenisty, N. & Ben-David, U. Quality control: Genome maintenance in pluripotent stem cells. *J. Cell Biol.* **204**, 153–163 (2014).
383. Bar, S. & Benvenisty, N. Epigenetic aberrations in human pluripotent stem cells. *EMBO J.* **38**, e101033 (2019).

384. Yan, Y.-W., Qian, E. S., Woodard, L. E. & Bejoy, J. Neural lineage differentiation of human pluripotent stem cells: Advances in disease modeling. *World J. Stem Cells* **15**, 530–547 (2023).
385. Yan, Y.-W., Qian, E. S., Woodard, L. E. & Bejoy, J. Neural lineage differentiation of human pluripotent stem cells: Advances in disease modeling. *World J. Stem Cells* **15**, 530–547 (2023).
386. Miura, Y. *et al.* Engineering brain assembloids to interrogate human neural circuits. *Nat. Protoc.* **17**, 15–35 (2022).
387. Amin, N. D. & Paşca, S. P. Mouse embryo models built from stem cells take shape in a dish. *Nature* **610**, 39–40 (2022).

SUPPLEMENTARY INDICES: FIGURES AND TABLES

FIGURE INDEX¹

Figure 1. Worldwide ASD prevalence in 2019. Dark green: 0-0.4 %; light green: 0.4-0.5 %; lavender: 0.5-0.6 %; violet: 0.6-0.7 %; dark violet: 0.7-1 %; grey: no data available. The figure was created by the author using R software (R Core Team, Vienna, Austria) to represent open access data, previously standardized by age and sex ^{7,8,10}	34
Figure 2. Number of genes linked to ASD according to the SFARI gene database (April 2023 release) ³² , distributed by their ASD-association score.	37
Figure 3. Protein coding genes associated with ASD according to the SFARI gene database (April 2023 release) ³² . Genes are distributed according to the chromosome where they are located. On the left, the total number of genes is indicated. On the right, the number of genes linked to ASD was corrected by the total number of protein coding genes on each chromosome.	37
Figure 4. DNA repair pathways in eukaryotic cells to correct double-strand breaks (DSBs): non-homologous end joining (NHEJ, left) and homology-directed repair (HDR, right). Image created with Biorender.	41
Figure 5. Genome editing systems based on protein-DNA recognition. A) Meganucleases. B) ZFNs. C) TALENs. Image created with Biorender.	41
Figure 6. Genomic editing systems based on CRISPR-Cas9. A) CRISPR-Cas9 systems; B) CRISPR interference (CRISPRi) for the repression of gene expression. C) CRISPR activator (CRISPRa) for the enhanced expression of genes. D) Adenine base editor (ABE) for the conversion of adenines into inosines. E) Cytosine base editor (CBE) for the conversion of cytosines into uridine. F) Prime editing system. Image created with Biorender.	43
Figure 7. Zebrafish developmental stages. Embryogenesis phases: zygote, cleavage, blastula, gastrula, segmentation, pharyngula and hatching. Image created with Biorender.	48
Figure 8. Gene expression silencing (knockdown) mediated by morpholinos. Image created with Biorender. ...	50
Figure 9. DNA oligonucleotide structure (A) compared to morpholinos (B). Image created with Biorender.	51
Figure 10. <i>BTBD8</i> expression pattern in different tissues and cell types. A) Global <i>BTBD8</i> protein expression across different human tissues. B) <i>BTBD8</i> RNA expression across different human brain regions. C) <i>BTBD8</i> RNA expression at the single cell level in human brain. Expression levels indicated in normalized transcript per million (nTPM) Images obtained from The Human Protein Atlas database (Creative Commons Attribution-ShareAlike 3.0 International License) ²⁰⁷	57
Figure 11. Protein structure of <i>BTBD8</i> main isoform. A) Schematic representation of the protein domains in <i>BTBD8</i> (representation created by the author). B) <i>BTBD8</i> predicted 3D structure by AlphaFold Monomer v2.0 pipeline (Creative Commons Attribution 4.0 license) ^{214,215}	58
Figure 12. Protein structure of <i>FOSL2</i> main isoform (representation created by the author). A) Schematic representation of the protein domains in <i>FOSL2</i> . B) <i>FOSL2</i> predicted 3D structure by AlphaFold Monomer v2.0 pipeline (Creative Commons Attribution 4.0 license) ^{214,215}	59
Figure 13. <i>FOSL2</i> expression pattern in different tissues and cell types. A) Global <i>FOSL2</i> protein expression across different human tissues. B) <i>FOSL2</i> RNA expression across different human brain regions. C) <i>FOSL2</i> RNA expression at the single cell level in human brain. Expression levels indicated in normalized transcript per million (nTPM) Images obtained from The Human Protein Atlas database (Creative Commons Attribution-ShareAlike 3.0 International License) ²⁰⁷	60
Figure 14. <i>TAOK1</i> expression pattern in different tissues and cell types. A) Global <i>TAOK1</i> protein expression across different human tissues. B) <i>TAOK1</i> RNA expression across different human brain regions. C) <i>TAOK1</i> RNA expression at the single cell level in human brain. Expression levels indicated in normalized transcript per million (nTPM).	

¹ Figures included in this doctoral thesis have been created by the author, Sara Veiga Rúa, except for Figures 10 to 15. These figures contain graphs available in public databases under a Creative Commons license.

Images obtained from The Human Protein Atlas database (Creative Commons Attribution-ShareAlike 3.0 International License) ²⁰⁷	62
Figure 15. Protein structure of TAOK1 main isoform. A) Schematic representation of the protein domains in TAOK1 (image created by the author). B) TAOK1 predicted 3D structure by AlphaFold Monomer v2.0 pipeline (Creative Commons Attribution 4.0 license) ^{214,215}	62
Figure 16. TAOK1 role in the MAPK and Salvador-Warts-Hippo pathways. Image created with Biorender.	63
Figure 17. hiPSCs in culture. A) Expected morphology of hiPSCs colonies. B) hiPSCs culture showing differentiation spots (red arrows) mixed with colonies with an appropriate shape (blue arrows).	74
Figure 18. Gene knockout strategy followed in this thesis. Image created with Biorender.	75
Figure 19. pSpCas9(BB)-2A-Puro (pX459) V2.0 schematic representation (Addgene #62988) ⁵⁰ . Image created by the author.	76
Figure 20. Oligonucleotides designed to be compatible with cloning into SpCas9 (BB)-2A-Puro (pX459) V2.0. In orange, compatible ends. N represent a generic 20 nucleotide sgRNA sequence.	76
Figure 21. Steps require to transform plasmids into E. coli. Image created with Biorender.	77
Figure 22. Guide efficiency analysis methods used in this thesis. Image created with Biorender.	79
Figure 23. Program used for PCR with hiPSCs samples using Phusion polymerase.	80
Figure 24. hiPSC colony grown from a single cell.	81
Figure 25. Program used for RT-qPCR with SYBR Green chemistry in the Quant Studio 5 (Applied Biosystems).	84
Figure 26. Program used for PCR with hiPSCs samples using AmpliTaq DNA polymerase. Image created using BioRender.com.	86
Figure 27. Schematic representation of the differentiation protocol used for the generation of cortical-like neurons, derived from Shi <i>et al.</i> ¹⁰⁵ Image created with Biorender.com.	87
Figure 28. Neural rosettes formed <i>in vitro</i> after 19 days of cortical differentiation.	88
Figure 29. Schematic representation of the differentiation protocol used for the generation of human brain organoids, as described by Velasco <i>et al.</i> ²⁸⁶ . CDM: cortical differentiation medium. Image created with Biorender.	89
Figure 30. Representation of the plasmid pLenti CMV GFP Neo (657-2), Addgene #17447. Figure generated by the author.	91
Figure 31. Workflow of a scRNA-seq experiment: from single cell obtention to data analysis. Image created with Biorender.	94
Figure 32. Experimental design of the scRNA-seq conducted on this thesis. Image created with Biorender.	95
Figure 33. Structure of the Single Cell 3' Gene Expression libraries constructed for scRNAseq. The constructs start and end with Illumina-P5 and Illumina-P7 sequences, respectively. They also contain 10 bp sample index sequences (i5 and i7) and a poly(dT) site used to capture mRNAs in the reverse transcription step. TruSeq Read 1 and 2 are standard Illumina sequencing primer sites. TruSeq Read 1 is used to sequence the 10x Barcodes (16 bp) and unique molecule identifier (UMI, 12 bp), while TruSeq Read 2 is used to sequence the cDNA insert.	96
Figure 34. Library qualitative analysis prior to scRNA-seq (Agilent TapeStation High Sensitivity D1000 ScreenTape, TapeStation Analysis Software 4.1.1). Libraries were diluted 1:10 for this analysis. A) High Sensitivity D1000 Ladder. B) Ctrl.BS-PBiPS37-Sv4F-1, day 48 organoids, replicate 1. C) Ctrl.BS-PBiPS37-Sv4F-1, day 48 organoids, replicate 2. D) Ctrl. BS PBiPS37-Sv4F-1, day 48 organoids, replicate 3. E) Ctrl.BS-PBiPS37-Sv4F-1, day 62 organoids, replicate 1. F) Ctrl.BS-PBiPS37-Sv4F-1, day 62 organoids, replicate 2. G) Ctrl.BS-PBiPS37-Sv4F-1, day 62 organoids, replicate 3. H) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 48 organoids, replicate 1. I) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 48 organoids, replicate 2. J) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 48 organoids, replicate 3. K) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 62 organoids, replicate 1. L) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 62 organoids, replicate 2. M) Ctrl.BS-PBiPS37-Sv4F-1-TAOK1-i2, day 62 organoids, replicate 3.	97
Figure 35. Morphometric analysis measurements conducted on zebrafish individuals: body length, cephalic size, and ocular size.	100
Figure 36. PCR conditions used in the amplification of CRISPR-Cas9 target sequences.	102
Figure 37. Schematic representation of the sgRNA synthesis for the knockout experiment in zebrafish. Image created with Biorender.com.	103
Figure 38. PCR program used to generate the sgRNA template for zebrafish knockout guides. Image created with Biorender.	104

Figure 39. PCR conditions used in the amplification of CRISPR-Cas9 target sequences.....	105
Figure 40. Thermocycler program used for Sanger sequencing with the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). Image created with Biorender.com.	106
Figure 41. Schematic representation of the sgRNAs designed for the knockout of <i>BTBD8</i> on hiPSCs using the CRISPR-Cas9 system. A) <i>BTBD8</i> _sgRNA_1. B) <i>BTBD8</i> _sgRNA_2. C) <i>BTBD8</i> _sgRNA_3. Image generated by the author using Synthego's Verify Guide Design webtool v1.3 ²⁸²	114
Figure 42. Representation of the plasmid pSpCas9 (BB)-2A-Puro (Addgene #62988) before and after the insertion of the guides to knockout <i>BTBD8</i> . A) Plasmid 62988 without modifications. B) Plasmid 62988 after the insertion of the <i>BTBD8</i> _sgRNA1. C) Plasmid 62988 after the insertion of the <i>BTBD8</i> _sgRNA2. D) Plasmid 62988 after the insertion of the <i>BTBD8</i> _sgRNA3.	115
Figure 43. Verification of the correct insertion of the knockout guides for <i>BTBD8</i> in the plasmid pSpCas9 (BB)-2A-Puro (Addgene #62988) by Sanger sequencing. A) Plasmids cloned with <i>BTBD8</i> _sgRNA1 (samples 1-6) aligned with the predicted sequence expected after the insertion (indicated as "Refseq" in the image). B) Plasmids cloned with <i>BTBD8</i> _sgRNA1 (samples 1-6) aligned with the predicted sequence expected after the insertion (indicated as "Refseq" in the image). C) Plasmids cloned with <i>BTBD8</i> _sgRNA1 (samples 1-6) aligned with the predicted sequence expected after the insertion (indicated as "Refseq" in the image). Black rectangles indicate the position of the inserted guide.	116
Figure 44. Puromycin kill curve for the cell line Ctrl.BS-PBiPS37-Sv4F-1. MTT viability assay performed after antibiotic treatment (24 h, 48 h, 72 h, 96 h, and 120 h). Data points represented as the mean $\pm$ SD.....	116
Figure 45. Alignment of the reference sequence of <i>BTBD8</i> (GRCh38), on top, with the alleles isolated by TOPO cloning from the bulk of edited cells. The target region of <i>BTBD8</i> _sgRNA1 is indicated in yellow.	117
Figure 46. Alignment of the reference sequence of <i>BTBD8</i> (GRCh38), on top, with the alleles isolated by TOPO cloning from the bulk of edited cells. The target region of <i>BTBD8</i> _sgRNA2 is indicated in red.	118
Figure 47. Alignment of the reference sequence of <i>BTBD8</i> (GRCh38), on top, with the alleles isolated by TOPO cloning from the bulk of edited cells. The target region of <i>BTBD8</i> _sgRNA3 is indicated in green.....	118
Figure 48. Clone 5 isolated from the bulk of cells edited with <i>BTBD8</i> _sgRNA1. Alignment of the reference sequence of <i>BTBD8</i> (GRCh38) with the Sanger sequences obtained from the clone of interest.	119
Figure 49. Alignment of the reference sequence of <i>BTBD8</i> (GRCh38) with the Sanger sequences obtained from several clones of interest isolated from the bulk of cells edited with <i>BTBD8</i> _sgRNA1. A) Clone 22 with an apparent insertion of 1 nt in homozygosis. B) Clone 25 with an apparent insertion of 1 nt in homozygosis. C) Clone 26 with an apparent deletion of 11 nts in homozygosis. D) Clone 28 with an apparent insertion of 1 nt in homozygosis.	119
Figure 50. Alignment with the reference sequence of <i>BTBD8</i> (GRCh38) with the Sanger sequences obtained from several clones of interest after TOPO cloning. A) Alleles present in the clone 39 isolated from the bulk of cells edited with <i>BTBD8</i> _sgRNA1. B) Alleles present in the clone 40 isolated from the bulk of cells edited with <i>BTBD8</i> _sgRNA1.	119
Figure 51. Predicted structures of the proteins generated by the mutant <i>BTBD8</i> alleles. A) Protein NP_001363060.1:p.Thr697del, encoded by the mutant <i>BTBD8</i> allele lacking 3 nucleotides (NC_000001.11:g.92177281_92177283delCAC). B) Protein NP_001363060.1:p.Leu696Thrfs*1, encoded by the mutant <i>BTBD8</i> allele lacking 10 nucleotides (NC_000001.11:g.92177279_92177288delCTCACAGGAA). Image generated using AlphaFold Monomer v2.0 pipeline ^{214,215}	120
Figure 52. Karyotype of the cell line Ctrl.BS-PBiPs37-Sv4F-1- <i>BTBD8</i> -i2.	120
Figure 53. Standard curves for the efficiency analysis of RT-qPCR primer pairs. A) <i>GAPDH</i> . B) <i>SOX2</i> . C) <i>NANOG</i> . D) <i>POU5F1</i>	121
Figure 54. RT-qPCR for pluripotency markers in the cell line Ctrl.BS-PBiPs37-Sv4F-1- <i>BTBD8</i> -i2 compared to the isogenic wild type cell line Ctrl.BS-PBiPs37-Sv4F-1. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.....	122
Figure 55. Quantification of immunofluorescence images for the pluripotency markers <i>SOX2</i> , <i>NANOG</i> , and <i>TRA-1-81</i> . Data represented as mean $\pm$ SD.	122
Figure 56. Immunofluorescence of the pluripotency marker <i>NANOG</i> in the cell line Ctrl.BS-PBiPs37-Sv4F-1- <i>BTBD8</i> -i2. Nuclear staining was performed with Hoechst.....	123
Figure 57. Immunofluorescence of the pluripotency markers <i>SOX2</i> and <i>TRA-1-81</i> in the cell line Ctrl.BS-PBiPs37-Sv4F-1- <i>BTBD8</i> -i2. Nuclear staining was performed with Hoechst.	124
Figure 58. Brightfield images of Ctrl.BS-PBiPs37-Sv4F-1- <i>BTBD8</i> -i2 edited cells and their isogenic control (Ctrl.BS-PBiPs37-Sv4F-1) taken during the three germ-layer differentiation protocol.	125

Figure 59. Standard curves for the efficiency analysis of RT-qPCR primer pairs. A) <i>AFP</i> . B) <i>NES</i> . C) <i>NCAM1</i>	125
Figure 60. RT-qPCR for germ-layer markers in the cell line Ctrl.BS-PBiPS37_Sv4F-1_ <i>BTBD8</i> -i2 comparison to the isogenic wild type cell line Ctrl.BS-PBiPS37_Sv4F-1. <i>NCAM1</i> expression was tested in mesoderm differentiated cells. <i>NES</i> expression was tested in ectoderm differentiated cells. <i>AFP</i> expression was tested in endoderm differentiated cells.....	126
Figure 61. Analysis of the presence of editing events in off-target positions for the guide <i>BTBD8</i> _sgRNA1. The region of each potential off-target was sequenced in the edited cell line Ctrl.BS-PBiPS37-Sv4F-1- <i>BTBD8</i> -i2 (indicated as “MT” in the image), and on the isogenic control Ctrl.BS-PBiPS37-Sv4F-1 (indicated as “WT” in the image). Both sequences were aligned with the genomic reference sequence, previously indicated in Table 30. Black rectangles indicate the predicted off-target region. Image created by the author in Codon Code Aligner v10.0.2.	128
Figure 62. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and <i>BTBD8</i> -mutant (Ctrl.BS-PBiPS37-Sv4F-1- <i>BTBD8</i> -i2) hiPSCs lines. A) Days 3-15 of the protocol. B) Days 18-35 of the protocol. Area is represented in logarithm scale as data show a lognormal distribution. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.	129
Figure 63. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and <i>BTBD8</i> -mutant (Ctrl.BS-PBiPS37-Sv4F-1- <i>BTBD8</i> -i2) hiPSCs lines on the first 18 days of the protocol. The same well was imaged per cell line on each time point.	129
Figure 64. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and <i>BTBD8</i> -mutant (Ctrl.BS-PBiPS37-Sv4F-1- <i>BTBD8</i> -i2) hiPSCs lines from day 18 to 56 of the protocol.	130
Figure 65. Immunostaining of the antigen KIAA1107 in brain organoids derived from the edited cell line Ctrl.BS-PBiPS37-Sv4F-1- <i>BTBD8</i> -i2 and their isogenic wild type control (Ctrl.BS-PBiPS37-Sv4F-1). Nuclear staining was performed with Hoechst.....	131
Figure 66. Characterization of the expression of cell-markers SOX2, PAX6, and MAP2 by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and <i>BTBD8</i> -mutant (Ctrl.BS-PBiPS37-Sv4F-1- <i>BTBD8</i> -i2) cell lines. Nuclear staining was performed with Hoechst.....	132
Figure 67. Characterization of the expression of cell-markers DLX2 and GAD67 by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1- <i>BTBD8</i> -i2) cell lines. Nuclear staining was performed with Hoechst.	132
Figure 68. Quantification of the expression the cell-markers SOX2, PAX6, MAP2, DLX2, and GAD67 in human brain organoids. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and <i>BTBD8</i> -mutant (Ctrl.BS-PBiPS37-Sv4F-1- <i>BTBD8</i> -i2) cell lines. Data represented as mean $\pm$ SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.	133
Figure 69. Characterization of the expression of cell-markers CTIP2 and SATB2 by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1- <i>BTBD8</i> -i2) cell lines. Nuclear staining was performed with Hoechst.	133
Figure 70. Characterization of the expression of cell-markers SOX2, PAX6, and MAP2 in cortical-like neurons derived from the Ctrl.BS-PBiPS37-Sv4F-1. Nuclear staining was performed with Hoechst.	134
Figure 71. Characterization of the expression of cell-markers CTIP2/BCL11B and SATB2 in cortical-like neurons derived from the Ctrl.BS-PBiPS37-Sv4F-1. Nuclear staining was performed with Hoechst.	135
Figure 72. Characterization of the expression of cell-markers BTBD8 and VGLUT1 in cortical-like neurons derived from the Ctrl.BS-PBiPS37-Sv4F-1. Nuclear staining was performed with Hoechst.....	135
Figure 73. Differentiation of cortical neurons from the cell lines Ctrl.BS-PBiPS37-Sv4F-1 and Ctrl.BS-PBiPS37-Sv4F-1- <i>BTBD8</i> -i2. Nuclear staining was performed with Hoechst.	136
Figure 74. Morphometric analysis of neurons in monolayer derived from the cell line Ctrl.BS-PBiPS37-Sv4F-1- <i>BTBD8</i> -i2 and its isogenic control Ctrl.BS-PBiPS37-Sv4F-1. A) Comparison of soma sizes. B) Comparison of the number of dendrites on each neuron. C) Comparison of the number of branches on each neuron. Data represented as mean + SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001.	136
Figure 75. Representative images of GFP-transfected neurons in monolayer.	137
Figure 76. Venn diagram of the number of proteins detected on each sample. False discovery rate 1 %.	137
Figure 77. Expression levels of BTBD8 in the proteomic assay. Normalization data was obtained taking Day 40 Ctrl.BS-PBiPS37-Sv4F as the reference sample. Quantification was performed in Scaffold v5.3.1. Red: day 40 cortical neurons, Ctrl.BS-PBiPS37-Sv4F; Blue: day 60 cortical neurons, Ctrl.BS-PBiPS37-Sv4F; Green: day 40 cortical neurons, Ctrl.BS-PBiPS37-Sv4F- <i>BTBD8</i> -i2; Yellow: day 60 cortical neurons, Ctrl.BS-PBiPS37-Sv4F- <i>BTBD8</i> -i2.	138

Figure 78. Expression levels of cell-type markers of interest in the proteomic assay. Normalization data was obtained taking Day 40 Ctrl.BS-PBiPS37-SvF4 as the reference sample. Quantification was performed in Scaffold v5.3.1. Red: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Blue: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Green: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4-*BTBD8-i2*; Yellow: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4-*BTBD8-i2*. 138

Figure 79. Expression levels of *BTBD8*-associated proteins implicated in endo/exocytosis pathways. Normalization data was obtained taking Day 40 Ctrl.BS-PBiPS37-SvF4 as the reference sample. Quantification was performed in Scaffold v5.3.1. Red: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Blue: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Green: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4-*BTBD8-i2*; Yellow: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4-*BTBD8-i2*. 139

Figure 80. Expression levels of *BTBD8*-associated proteins of interest. Normalization data was obtained taking Day 40 Ctrl.BS-PBiPS37-SvF4 as the reference sample. Quantification was performed in Scaffold v5.3.1. Red: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Blue: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4; Green: day 40 cortical neurons, Ctrl.BS-PBiPS37-SvF4-*BTBD8-i2*; Yellow: day 60 cortical neurons, Ctrl.BS-PBiPS37-SvF4-*BTBD8-i2*. 141

Figure 81. Expected outcome of the knockdown with the splice blocking morpholino that targets the first splice donor site. A) Splicing of the *fosl2* pre-mRNA under normal circumstances. B) Inclusion of intron 1 in the *fosl2* pre-mRNA due to the action of the splice blocking morpholino. Coding regions within exons are represented in orange. The target region for the morpholino is represented in yellow. Green and blue dots indicate the positions on pre-mRNA of the splice junctions, which are recognized by the spliceosome for correct processing. 143

Figure 82. Agarose gel (1.5 %) loaded with the amplicons obtained after RT-PCR of cDNA samples from wild type control zebrafish and *fosl2* morphant individuals. 143

Figure 83. Bright-field images of 48 hpf zebrafish individuals. From left to right: control, treated with MO-*fosl2*, treated with both MO-*fosl2* and MO-*p53*. 144

Figure 84. Morphometric analysis of the individuals treated with morpholines at 48 hpf. Analysis conducted with wild type individuals (control), individuals injected with *fosl2* morpholine (MO-*fosl2*) and individuals co-injected with *fosl2* and *p53* morpholines (MO-*fosl2* + MO-*p53*). A) Full body length. B) Cephalic area. C) Cephalic diameter. D) Ocular area. E) Ocular diameter. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 145

Figure 85. Morphometric analysis of the individuals treated with morpholines at 48 hpf. Analysis conducted with wild type individuals (control), individuals injected with *fosl2* morpholine (MO-*fosl2*) and individuals co-injected with *fosl2* and *p53* morpholines (MO-*fosl2* + MO-*p53*). A) Ratio between the cephalic diameter and the full body length of the individuals. A) Ratio between the ocular diameter and the cephalic diameter of the individuals. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 146

Figure 86. Cartilage staining of *fosl2* morphants. A) Ceratohyal angle. B) Palatoquadrate length. Data represented as 10-90 percentile. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 147

Figure 87. General representation of the movements of the individuals at 120 hpf in light and dark cycles. Analysis conducted with wild type individuals (control), individuals injected with *fosl2* morpholine (MO-*fosl2*) and individuals co-injected with *fosl2* and *p53* morpholines (MO-*fosl2* + MO-*p53*). A) Movements detected in light cycles. B) Movements detected in dark cycles. Data represented as mean + SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 147

Figure 88. Light-dark response test of *fosl2* morphant zebrafish. Analysis conducted with wild type individuals (control), individuals injected with *fosl2* morpholine (MO-*fosl2*) and individuals co-injected with *fosl2* and *p53* morpholines (MO-*fosl2* + MO-*p53*). Movements were measured as distance (in pixels) travelled per minute. Data represented as mean + SD. 148

Figure 89. Schematic representation of the sgRNAs designed for the knockout of *TAOK1* on hiPSCs using the CRISPR-Cas9 system. A) *TAOK1_sgRNA_1*. B) *TAOK1_sgRNA_2*. C) *TAOK1_sgRNA_3*. Image generated by the author using Synthego's Verify Guide Design webtool v1.3²⁸². 149

Figure 90. Representation of the plasmid pSpCas9 (BB)-2A-Puro (Addgene #62988) before and after the insertion of the guides to knockout *BTBD8*. A) Plasmid 62988 without modifications. B) Plasmid 62988 after the insertion of the *BTBD8_sgRNA1*. C) Plasmid 62988 after the insertion of the *BTBD8_sgRNA2*. D) Plasmid 62988 after the insertion of the *BTBD8_sgRNA3*. Images generated by the author using Benchling webtool. 150

Figure 91. Verification of the correct insertion of the knockout guides for *TAOK1* in the plasmid pSpCas9 (BB)-2A-Puro (Addgene #62988) by Sanger sequencing. A) Plasmids cloned with *TAOK1_sgRNA1* (samples 1-6) aligned with

- the predicted sequence expected after the insertion (indicated as “Refseq” in the image). B) Plasmids cloned with *TAOK1_sgRNA1* (samples 1-6) aligned with the predicted sequence expected after the insertion (indicated as “Refseq” in the image). C) Plasmids cloned with *TAOK1_sgRNA1* (samples 1-6) aligned with the predicted sequence expected after the insertion (indicated as “Refseq” in the image). Black rectangles indicate the position of the inserted guide. Image created by the author in Codon Code Aligner v10.0.2. 151
- Figure 92. Traces tab of the ICE analysis conducted for the guides designed to knockout *TAOK1*. Sanger sequencing data of the edited pool of cells is compared to the wild type isogenic control for the three different guides. A) *TAOK1_sgRNA1*. B) *TAOK1_sgRNA2*. C) *TAOK1_sgRNA3*. The target sequence for each guide is underlined in black, while the PAM sequence is indicated underlined with red dots. 152
- Figure 93. ICE analysis results for the knockout guides for *TAOK1*. Disagreement plot (left) with the wild type sequence in orange and the edited samples in green. Histogram of the frequencies of indels in the edited samples (right). R^2 (Pearson correlation coefficient) indicates the model fit value. A) *TAOK1_sgRNA1*, R^2 0.88, editing efficiency 76 %. B) *TAOK1_sgRNA2*, R^2 0.97, editing efficiency 89 %. C) *TAOK1_sgRNA3* R^2 0.93, editing efficiency 80%. 153
- Figure 94. Predicted efficiency of the guide *TAOK1_sgRNA1* (in yellow). Alignment of the reference sequence of *TAOK1* (GRCh38), on top, with the alleles isolated by TOPO cloning from the bulk of edited cells. Image created by the author using the SnapGene software. 154
- Figure 95. Clone 24 isolated from the edition with *TAOK1_sgRNA1*. Alignment with the reference sequence of *TAOK1* (GRCh38) with the Sanger sequences obtained from the clone of interest. 155
- Figure 96. Clone 10 isolated from the edition with *TAOK1_sgRNA1*. Alignment with the reference sequence of *TAOK1* (GRCh38) with the Sanger sequences obtained from the clone of interest. 155
- Figure 97. Clone 18 isolated from the edition with *TAOK1_sgRNA1*. Alignment with the reference sequence of *TAOK1* (GRCh38) with the Sanger sequences obtained from the clone of interest. 155
- Figure 98. Clone 28 isolated from the edition with *TAOK1_sgRNA1*. Alignment with the reference sequence of *TAOK1* (GRCh38) with the Sanger sequences obtained from the clone of interest. 155
- Figure 99. Alignment with the reference sequence of *TAOK1* (GRCh38) with the Sanger sequences obtained from several clones of interest after TOPO cloning. A) Alleles present in the clone 4 isolated from the edition with *TAOK1_sgRNA1*. B) Alleles present in the clone 6 isolated from the edition with *TAOK1_sgRNA1*. A) Alleles present in the clone 7 isolated from the edition with *TAOK1_sgRNA1*. A) Alleles present in the clone 8 isolated from the edition with *TAOK1_sgRNA1*. 155
- Figure 100. Predicted structures of the proteins generated by the mutant *TAOK1* alleles. A) Protein NP_065842.1:p.Gly35Alafs*48 derived from NC_000017.11:g.29451652delG. B) Protein NP_065842.1:p.His36Lysfs*14 derived from NC_000017.11:g.29451653_29451657delCCATG. Image generated using AlphaFold Monomer v2.0 pipeline^{214,215}. 156
- Figure 101. Karyotype of the cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2*. 157
- Figure 102. RT-qPCR for pluripotency markers in the knockout cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2* compared to the isogenic wild type cell line Ctrl.BS-PBiPs37-Sv4F-1. Data represented as the geometric mean of the fold change and the geometric SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 157
- Figure 103. Quantification of immunofluorescence images for the pluripotency markers SOX2, NANOG, and TRA-1-81. 158
- Figure 104. Immunofluorescence staining of the pluripotency marker NANOG in the cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2*. Nuclear staining was conducted with Hoechst. 159
- Figure 105. Immunofluorescence staining of the pluripotency marker SOX2 in the cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2*. Nuclear staining was conducted with Hoechst. 160
- Figure 106. Immunofluorescence staining of the pluripotency marker TRA-1-81 in the cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2*. Nuclear staining was conducted with Hoechst. 161
- Figure 107. Brightfield images of the three germ-layer differentiation conducted using Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2* edited cells and their isogenic control (Ctrl.BS-PBiPs37-Sv4F-1). 162
- Figure 108. RT-qPCR for germ layer markers in the cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2* compared to the isogenic wild type cell line Ctrl.BS-PBiPs37-Sv4F-1. *NCAM1* expression was tested in mesoderm differentiated cells. *NES* expression was tested in ectoderm differentiated cells. *AFP* expression was tested in endoderm differentiated cells. 162
- Figure 109. Analysis of the presence of editing events in off-target positions for the guide *TAOK1_sgRNA1*. The region of each potential off-target was sequenced in the edited cell line Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2* (indicated as “MT” in the image), and on the isogenic control Ctrl.BS-PBiPs37-Sv4F-1 (indicated as “WT” in the

- image). Both sequences were aligned with the genomic reference sequence, previously indicated in Table 37. Black rectangles indicate the predicted off-target region. Image created by the author in Codon Code Aligner v10.0.2. 165
- Figure 110. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and *TAOK1*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2*) hiPSCs lines on the first 18 days of the protocol. The same well was imaged per cell line on each time point. 165
- Figure 111. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and *TAOK1*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2*) hiPSCs lines from day 18 to 56 of the protocol. 165
- Figure 112. Comparison of the size of human brain organoids derived from the wild type (Ctrl.BS-PBiPS37-Sv4F-1) and *TAOK1*-mutant (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2*) hiPSCs lines. A) Days 3-15 of the protocol. B) Days 18-42 of the protocol. Area is represented in logarithm scale as data show a lognormal distribution. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 166
- Figure 113. Immunofluorescence for the antigen *TAOK1* in human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and edited (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2*) cell lines. 167
- Figure 114. Characterization of the expression of cell-markers *SOX2*, *PAX6*, and *MAP2* by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2*) cell lines. 167
- Figure 115. Characterization of the expression of cell-markers *DLX2* and *GAD67* by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2*) cell lines. 168
- Figure 116. Quantification of the expression the cell-markers *SOX2*, *PAX6*, and *MAP2* in human brain organoids. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2*) cell lines. Data represented as mean $\pm$ SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 168
- Figure 117. Characterization of the expression of cell-markers *CTIP2/BCL11B* and *SATB2* by immunofluorescence. Human brain organoids derived from the WT (Ctrl.BS-PBiPS37-Sv4F-1) and KO (Ctrl.BS-PBiPS37-Sv4F-1-*TAOK1-i2*) cell lines. 169
- Figure 118. UMAP representations of scRNA-seq data integrated by cell line: Ctrl.BS-PBiPs37-Sv4F-1 (A, C, E) and Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2* (B, D, F). A-B) Clusters identified on each group of samples with integrated information of both time point (day 48, day 62). C-D) Clusters identified on each group of samples with separated representations for both time points (day 48, day 62). E-F) Day 48 and day 62 distribution in the UMAP. 170
- Figure 119. Clusters identified after integrating scRNA-seq Ctrl.BS-PBiPS37-SvF4-1 samples. The number of cells from each time point analysed (day 48, day 62) is represented for each cluster. Day 48 samples: dark grey; day 62 samples: light grey. 171
- Figure 120. Clusters identified after integrating scRNA-seq Ctrl.BS-PBiPS37-SvF4-1-*TAOK1-i2* samples. The number of cells from each time point analysed (day 48, day 62) is represented for each cluster. Day 48 samples: dark grey; day 62 samples: light grey. 172
- Figure 121. UMAP representation of scRNA-seq data integrated by time: day 48 samples (A, C, E) and day 62 samples (B, D, F). A-B) Clusters identified on each group of samples with integrated information of both cell lines (Ctrl.BS-PBiPs37-Sv4F-1, Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2*). C-D) Clusters identified on each group of samples with separated representations for both cell lines (Ctrl.BS-PBiPs37-Sv4F-1 identified as "WT", Ctrl.BS-PBiPs37-Sv4F-1-*TAOK1-i2* identified as "mutant"). E-F) Cell line distribution in the UMAP. 174
- Figure 122. Clusters identified after integrating scRNA-seq day 48 samples. The number of cells from each cell line (Ctrl.BS-PBiPS37-SvF4-1, Ctrl.BS-PBiPS37-SvF4-1-*TAOK1-i2*) is represented for each cluster. Ctrl.BS-PBiPS37-SvF4-1 samples: dark grey; Ctrl.BS-PBiPS37-SvF4-1-*TAOK1-i2* samples: light grey. 175
- Figure 123. Clusters identified after integrating scRNA-seq day 62 samples. The number of cells from each cell line (Ctrl.BS-PBiPS37-SvF4-1, Ctrl.BS-PBiPS37-SvF4-1-*TAOK1-i2*) is represented for each cluster. Ctrl.BS-PBiPS37-SvF4-1 samples: dark grey; Ctrl.BS-PBiPS37-SvF4-1-*TAOK1-i2* samples: light grey. 175
- Figure 124. Alignment of the proteins *TAOK1* (above) and *taok1* (below) generated by the Clustal Omega program (EMBL-EBI) ³⁰¹. In orange it is marked the protein kinase domain of the protein (28 - 281 aa). 177
- Figure 125. Schematic representation of the sgRNAs designed for the knockout of *taok1* in zebrafish using the CRISPR-Cas9 system. A) *taok1_sgRNA1* (target located in exon 4). B) *taok1_sgRNA2* (target located in exon 6). Image generated by the author using Synthego's Verify Guide Design webtool v1.3²⁸². 178
- Figure 126. Sequences of interest for the generation of sgRNAs by *in vitro* transcription. A) Single-strand DNA oligonucleotides with the target sequences (crRNA) selected for knocking out *taok1*. They also include the T7

promoter sequence and an annealing region complementary to the tracrRNA DNA-template. B) Single-strand DNA oligonucleotide that encodes the tracrRNA portion of a sgRNA for CRISPR-Cas9. C) Full length DNA templates (117 bp) of the designed sgRNAs for the knockout of *taok1*..... 179

Figure 127. Synthesis verification of sgRNA DNA-templates (117 bp) in an agarose gel loaded with a 100 bp DNA ladder. 179

Figure 128. Wild type sequence of zebrafish in the target region for the selected knockout guides (in black boxes). A) Target region for *taok1*_sgRNA1. B) Target region for *taok1*_sgRNA2. 179

Figure 129. Heteroduplex analysis conducted in an acrylamide gel to verify guide efficiency. L: ladder 100 bp. P: pool of injected embryos from the F0. WT: pool of wild type non-injected embryos. 180

Figure 130. Heteroduplex analysis conducted in an acrylamide gel to genotype adult zebrafish from the F0 generation (n=20). Lanes 1 and 16 were loaded with 100 bp ladder, the remaining wells were loaded with amplicons from the target region of *taok1*_sgRNA2 obtained from different F0 individuals (ordered from 1 to 20, starting from the right). 180

Figure 131. Heteroduplex analysis conducted in an acrylamide gel to genotype adult zebrafish from the F1 generation (n=7). A 100 bp ladder was loaded in lane 1, the remaining wells were loaded with amplicons from the target region of *taok1*_sgRNA2 obtained from different F1 individuals (ordered from 1 to 7, starting from the right). 181

Figure 132. Alleles identified through bacterial cloning in the F1 generation. Sequences were aligned with the reference sequence of *taok1* exon 6 (indicated as Refseq *taok1*), the target region of *taok1*_sgRNA2. One of the sequences carries an insertion of 4 bp (in the middle). The new stop codon is marked in black. 181

Figure 133. Representation of the types of electropherograms obtained after fragment analysis for the genetic characterization of zebrafish individuals (ladder in orange, samples in green). A) Homozygous individual with two wild type alleles (415 bp). B) Heterozygous individual with one wild type (415 bp) and one edited (419 bp) allele. C) Homozygous individual with two edited alleles (419 bp). 182

Figure 134. Representative images of individuals from the F2 generation at 48 hfp and 96 hfp, classified by their genotype. 183

Figure 135. Morphometric analysis of the F2 generation at 48 hfp. Analysis conducted with wild type individuals (WT ^{415/415}), heterozygous individuals that carry the knockout mutation in one allele (HT ^{415/419}) and homozygous individuals that carry the mutations in both alleles (KO ^{419/419}). A) Full body length. B) Cephalic diameter. C) Ocular diameter. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 183

Figure 136. Morphometric analysis of the F2 generation at 48 hfp. Analysis conducted with wild type individuals (WT ^{415/415}), heterozygous individuals that carry the knockout mutation in one allele (HT ^{415/419}) and homozygous individuals that carry the mutations in both alleles (KO ^{419/419}). A) Ratio between the ocular diameter and the cephalic diameter of the individuals. B) Ratio between the cephalic diameter and the full body length of the individuals. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 184

Figure 137. Morphometric analysis of the F2 generation at 96 hfp. Analysis conducted with wild type individuals (WT ^{415/415}) and their heterozygous siblings that carry the knockout mutation (HT ^{415/419}). A) Full body length. B) Cephalic diameter. C) Ocular diameter. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 185

Figure 138. Morphometric analysis of the F2 generation at 96 hfp. Analysis conducted with wild type individuals (WT ^{415/415}) and heterozygous individuals that carry the knockout mutation in one allele (HT ^{415/419}). A) Ratio between the ocular diameter and the cephalic diameter of the individuals. B) Ratio between the cephalic diameter and the full body length of the individuals. Data represented as 10-90 percentile. The p-value obtained after statistical analysis is indicated in the table below. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 186

Figure 139. Light-dark response of 96 hfp individuals from the F2 generation. Analysis conducted with wild type individuals (WT ^{415/415}) and heterozygous individuals that carry the knockout mutation in one allele (HT ^{415/419}). Movements were measured as distance (in pixels) travelled per minute. Data represented as mean + SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001. 186

Figure 140. General representation of the movements of the individuals from the F2 generation (96 hfp) in light and dark cycles. Analysis conducted with wild type individuals (WT ^{415/415}) and heterozygous individuals that carry

the knockout mutation in one allele (HT^{415/419}). A) Movements detected in light cycles. B) Movements detected in dark cycles. Data represented as mean + SD. Statistically significant p-values are represented with the symbol (*): * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001..... 187

Figure 141. Cellular components enriched (orange) and depleted (blue) in cortical neurons derived from the cell line Ctrl. BS PBiPS37-Sv4F-1-*BTBD8-i2* in comparison to the wild type isogenic control. Comparison between neuronal cultures from day 40 (A) and day 60 (B). Represented the top 10 results obtained with FunRich v3.1.3. 262

Figure 142. Biological processes enriched (orange) and depleted (blue) in cortical neurons derived from the cell line Ctrl. BS PBiPS37-Sv4F-1-*BTBD8-i2* in comparison to the wild type isogenic control. Comparison between neuronal cultures from day 40 (A) and day 60 (B). Represented the top 10 results obtained with FunRich v3.1.3. 263

Figure 143. Reactome pathways enriched (orange) and depleted (blue) in cortical neurons derived from the cell line Ctrl. BS PBiPS37-Sv4F-1-*BTBD8-i2* in comparison to the wild type isogenic control. Comparison between neuronal cultures from day 40 (A) and day 60 (B). Represented the top 10 results obtained with FunRich v3.1.3. 264

Figure 144. Molecular functions enriched (orange) and depleted (blue) in cortical neurons derived from the cell line Ctrl. BS PBiPS37-Sv4F-1-*BTBD8-i2* in comparison to the wild type isogenic control. Comparison between neuronal cultures from day 40 (A) and day 60 (B). Represented the top 10 results obtained with FunRich v3.1.3. 265

TABLE INDEX

Table 1. SFARI gene database scoring system criteria. Information obtained from SFARI gene database (April 2023 release) ³²	36
Table 2. Pharmacological treatments that have been tested to alleviate ASD symptoms.	39
Table 3. Non-pharmacological therapies that have been used to alleviate ASD symptoms.	40
Table 4. Stem cell classification attending to their differentiation potential.	44
Table 5. Delivery systems tested for reprogramming of somatic cells into iPSCs.	46
Table 6. Classification and characteristics of morpholinos.	51
Table 7. <i>TAOK1</i> mutations associated with developmental delay with or without intellectual impairment or behavioural abnormalities (MIM 619575).	64
Table 8. hiPSCs lines used in this thesis. Data was obtained from the Spanish National Bank of Cell Lines ²⁷⁸ . ..	73
Table 9. One-step digestion and ligation of guides into plasmid.	77
Table 10. Sequencing primers for the identification of genetic edition in hiPSCs.	80
Table 11. Antibodies used for the pluripotency analysis of the hiPSCs cell lines.	83
Table 12. cDNA synthesis reaction set-up.	83
Table 13. Sequences and amplicon size of the RT-qPCR primers selected for pluripotency analysis.	84
Table 14. RT-qPCR reaction set-up with SYBR Green chemistry.	84
Table 15. RT-qPCR primer pairs designed for the amplification of specific markers for each germ layer.	86
Table 16. Components of the neural maintenance medium (NMM) used for the differentiation of cortical-like neurons.	87
Table 17. Recipe for preparation of cortical differentiation medium 1 (CDMI).	89
Table 18. Recipe for Preparation of cortical differentiation medium 2 (CDMII).	89
Table 19. Preparation of cortical differentiation medium 3 (CDMIII).	90
Table 20. Preparation of cortical differentiation medium 4 (CDMIV).	90
Table 21. Antibodies used for the characterization of differentiated cell types.	92
Table 22. Viability of the single-cell dissociated organoids used for scRNA-seq.	95
Table 23. Sequences of the morpholinos used in this thesis.	99
Table 24. cDNA synthesis reaction set-up.	102
Table 25. Sequences and amplicon size of the RT-PCR primers.	102
Table 26. Relevant sequences used for the knockout guide design in zebrafish. The sequences indicated in blue are complementary.	103
Table 27. Sequence alterations identified by TOPO cloning of the cells treated with <i>BTBD8</i> _sgRNA1.	117
Table 28. Sequence alterations identified by TOPO cloning of the cells treated with <i>BTBD8</i> _sgRNA2.	118
Table 29. Sequence alterations identified by TOPO cloning of the cells treated with <i>BTBD8</i> _sgRNA3.	118
Table 30. Predicted off-target regions for <i>BTBD8</i> _sgRNA1 identified using the Cas-OFFinder ⁶⁸ and Synthego's Verify Guide Desing ²⁸² webtools. Mismatches between the off-target region and the sgRNA sequence are indicated in lowercase.	126
Table 31. Sequencing primers designed to study the off-target regions of the <i>BTBD8</i> _sgRNA1.	127
Table 32. Mortality rate 48 h after morpholine treatment in zebrafish embryos.	142
Table 33. Morphometric measurements conducted in the individuals treated with morpholines at 48 hpf.	144
Table 34. Morphometric ratios of the individuals treated with morpholines at 48 hpf.	146

Table 35. P-values ($\alpha = 0.05$) obtained in the statistical analyses conducted with data obtained from light-dark test. Statistically significant values are indicated with the symbol (*).	148
Table 36. Sequence alterations identified by TOPO cloning of the cells treated with <i>TAOK1</i> _sgRNA1.....	154
Table 37. Predicted off-target regions for <i>TAOK1</i> _sgRNA1 identified using the CasOFFinder and Synthego's Verify Guide Desing webtools. Mismatches between the off-target region and the sgRNA sequence are indicated in lowercase.	163
Table 38. Sequencing primers designed to study the off-target regions of the <i>TAOK1</i> _sgRNA1.....	163
Table 39. Sequencing primers for the identification of mutations in zebrafish in the target regions for the <i>taok1</i> knockout guides.	179
Table 40. Genotypes of the offsprings obtained after crossing F1 individuals.....	182
Table 41. Morphometric measurements of the F2 generation at 48 hfp.....	183
Table 42. Morphometric ratios of the individuals from the F2 generation at 48 hpf.	184
Table 43. Morphometric measurements of the F2 generation at 96 hfp.....	185
Table 44. Morphometric ratios of the individuals from the F2 generation at 96 hpf.	185

ANNEXES

ANNEXE I: APPROVAL OF THE ANIMAL EXPERIMENTATION ETHICS COMMITTEE.



VICERREITORADO DE INVESTIGACIÓN
E INNOVACIÓN
Oficina de Investigación e Tecnoloxía

Edificio CACTUS - Campus universitario sur
15782 Santiago de Compostela
Tel. 981 547 040 - Fax 981 547 077
Correo electrónico: calinfo@usc.es
<http://investi.usc.es>

Informe del Comité de Ética de Experimentación Animal (CEEA) de los centros usuarios de animales de experimentación de la USC en el Campus de Lugo

El CEEA de los centros usuarios de animales de experimentación de la USC en el Campus de Lugo, tras evaluar el Proyecto titulado "El pez cebra y su aplicación en biomedicina, acuicultura y medio ambiente" del que es Investigadora responsable D^a. Ana Quelle Regaldie, acordó con fecha 30 de enero de 2020 emitir

INFORME FAVORABLE

para la realización de dicho Proyecto, así como los procedimientos que incluye, en las instalaciones del establecimiento usuario Animalario de la Facultad de Veterinaria, con número de registro ES270280346401 (AE-LU-003), y siempre que, en cumplimiento del RD 53/2013, se obtenga la correspondiente autorización administrativa (en concreto, en relación con el artículo 27.2 y el apartado 1.b de la cláusula de salvaguardia enunciada en la disposición adicional segunda).

En Lugo, a 30 de enero de 2020

Fdo.: La presidenta

Sonia Vázquez Rodríguez

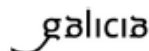
*Supervisora das instalacións de experimentación,
cria e subministro de animais de laboratorio da
Universidade de Santiago de Compostela*

Responsable administrativo:	Nombre: Montserrat Valcárcel Armesto Cargo: Vicerreitora de Coordinación do Campus de Lugo	VºBº Firma y sello
-----------------------------	---	------------------------

ANNEXE II: AUTHORIZATION OF ANIMAL EXPERIMENTATION PROJECTS.



Xefatura territorial de Lugo
Servizo de Gandería
Ronda da Muralla, 70
Lugo



RESOLUCIÓN DE AUTORIZACIÓN DE PROXECTOS DE EXPERIMENTACIÓN ANIMAL

Expediente núm.: 01/20/LU-003

Interesado: **Ana Quelle Regaldie**

Procedemento: RESOLUCIÓN DE AUTORIZACIÓN DE PROXECTOS DE EXPERIMENTACIÓN ANIMAL

Data de inicio: 1 de marzo de 2020

Forma de inicio: solicitude da interesada

ANTECEDENTES

A interesada, como responsable do proxecto “**El pez cebra y su aplicación en biomedicina, acuicultura y medio ambiente**” presentou con data 12 de febreiro de 2020 solicitude para a realización do proxecto de experimentación animal cuxos datos se detallan a continuación:

Denominación do proxecto : El pez cebra y su aplicación en biomedicina, acuicultura y medio ambiente

Nome do centro usuario: Animalario da Facultade de veterinaria(AE-LU-003).

Persoa responsable do proxecto: Ana Quelle Regaldie

Establecemento onde se realizarán os procedementos do proxecto (ou lugar xeográfico no caso de traballos de campo) Animalario da Facultade de veterinaria(AE-LU-003).

Clasificación do proxecto : Tipo III.

CONSIDERACIÓNS LEGAIS E TÉCNICAS

Real decreto 53/2013, de 1 de febreiro (BOE núm. 34, do 8 de febreiro), polo que se establecen as normas básicas aplicables para a protección dos animais utilizados en experimentación e outros fins científicos, incluíndo a docencia, establece no seu artigo 33 as condicións de autorizacións dos proxectos con animais de experimentación.

Artigo 88 da Lei 39/2015, do 1 de outubro, do procedemento administrativo común das administracións públicas (BOE núm. 236, do 2 de outubro) establece que a resolución que poña fin ao procedemento decidirá todas as cuestións formuladas polos interesados e aquelas outras derivadas del.

Esta xefatura territorial é competente para ditar resolución de conformidade co artigo 11 do Decreto 245/2009, do 30 de abril, polo que se regulan as delegacións territoriais da Xunta de Galicia e o Decreto 149/2018, do 5 de decembro, polo que establece a estrutura orgánica da Consellería do Medio Rural. O Servizo Provincial de Gandería de Lugo revisou a documentación presentada coa solicitude e visto o informe favorable da avaliación do proxecto, de data 30 de xaneiro de 2020 emitido polo órgano habilitado, e o Comité de Ética da Universidade de Santiago de Compostela, esta xefatura territorial resolve AUTORIZAR o proxecto solicitado.

CSL 2014-00001
Verificación: https://sede.xunta.gal/ove



Xacobeo 2021

A autorización deste proxecto terá unha duración de 5 ANOS e unha vez que transcorran deberán renova-la.

A autorización é unicamente válida nas condicións que figuran no expediente. Ante calquera cambio significativo no proxecto que poida ter efectos negativos sobre o benestar dos animais, deberá solicitar a confirmación da autorización ao Servizo Provincial de Gandería.

Esta autorización poderá ser suspendida no caso de que o proxecto non se leve a cabo de acordo coas condicións de autorización e retirala previo expediente tramitado ao que se lle dará audiencia.

Contra a presente resolución, que non pon fin á vía administrativa, poderá interpor recurso de alzada ante o conselleiro de Medio Rural da Xunta de Galicia no prazo dun mes contado a partir da recepción da notificación da presente resolución, conforme cos artigos 121 e 122 da Lei 39/2015, do 1 de outubro, do procedemento administrativo común das administracións públicas.

Documento asinado electronicamente á marxe

Asinado por: GLESIAS PONTAL, MARIA OLGA
Cargo: Xefa Territorial
Data e hora: 06/02/2020 13:46:44

CVE: 23MUCW0H1S
Verificación: <https://sede.xunta.gal/vea>

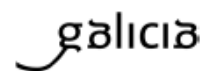


Xacobeo 2021

ANNEXE III: ANIMAL EXPERIMENTATION TRAINING CERTIFICATE.

 Instituto de Investigación y Formación Agraria y Pesquera CONSEJERÍA DE AGRICULTURA, PESCA Y DESARROLLO RURAL		 UNIÓN EUROPEA
CERTIFICADO DE CAPACITACIÓN		
Artículo 13, Orden ECC/566/2015, de 20 de marzo por la que se establecen los requisitos de capacitación que debe cumplir el personal que maneje animales utilizados, criados o suministrados con fines de experimentación y otros fines científicos, incluyendo la docencia.		
1. IDENTIFICACIÓN		
1.1. Apellidos y Nombre VEIGA RÚA, SARA		
1.2. DNI/NIE 78810873-Q		
1.3. Fecha de Nacimiento 17/09/1995	1.4. Lugar y país de nacimiento ORDES (ESPAÑA)	1.5. Nacionalidad ESPAÑOLA
2. FECHA DE RESOLUCIÓN DEL RECONOCIMIENTO DE LA CAPACITACIÓN INICIAL Y N° DE CERTIFICADO		
20/11/2018 - EXP-000882		
3. FUNCIONES Y GRUPOS DE ESPECIES ANIMALES PARA LOS QUE SE RECONOCE LA CAPACITACIÓN (Artículo 3 y Anexo II Orden ECC/566/2015, de 20 de marzo)		
a) Cuidado de los animales b) Eutanasia de los animales c) Realización de los procedimientos		
GRUPO DE ESPECIES ANIMALES		
8. Peces y anfibios		
4. ORGANISMO QUE EXPIDE EL CERTIFICADO		
4.1. Nombre y dirección del organismo que expide el certificado CONSEJERÍA DE AGRICULTURA, PESCA Y DESARROLLO RURAL INSTITUTO DE INVESTIGACIÓN Y FORMACIÓN AGRARIA Y PESQUERA Edificio de Servicios Múltiples "los Bermejales" Avda. de Grecia s/n. 41012 Sevilla (España)		
4.2. Teléfono 958 056 000	4.3. Fax 955 519 107	4.4. Correo electrónico experimentacion.ifapa@juntadeandalucia.es
4.5. Fecha de expedición* 20/11/2018	4.6. Lugar SEVILLA	4.7. Sello
4.8. Nombre y Firma		
El Presidente del Instituto de Investigación y Formación Agraria y Pesquera 		
Fdo. Jerónimo José Pérez Parra 		
* La fecha de expedición coincide con el inicio del periodo de vigencia establecido en el artículo 21.8 de la Orden ECC/566/2015, de 20 de marzo.		

ANNEXE IV: APPROVAL OF THE ETHICS COMMITTEE FOR THE PROJECT INCLUDING THE USE OF hiPSCs.



DICTAMEN DEL COMITÉ DE ÉTICA DE LA INVESTIGACIÓN DE SANTIAGO-LUGO

Ana Estany Gestal, Secretaria del Comité de Ética de la Investigación de Santiago-Lugo,

CERTIFICA:

Que este Comité evaluó en su reunión del día 19 de enero de 2021 el estudio:

Título: Enfoque integral y traslacional del diagnóstico y tratamiento de trastornos del neurodesarrollo (TEA, TDAH, DI) y del TOC: Una aproximación genética y funcional.

Versión: 3.0

Promotor/a: Ángel Carracedo Álvarez

Código de Registro: 2020/400

Y que este Comité, tomando en consideración la pertinencia del estudio, el conocimiento disponible, los requisitos legales aplicables y los Procedimientos Normalizados de Trabajo del Comité, emite un dictamen **FAVORABLE** para la realización del citado estudio.

Documento asinado dixitalmente por:
Ana Estany Gestal (27/01/2021 14:06)
<https://sede.xunta.gal/cve?dcve=SAC-C4C4-80A4-0F40-TG3H-L7YL-15M-W161-1752-8080-09>



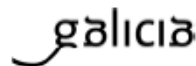
NOTA: Se le recuerda que en el caso de que en este estudio se recluten pacientes, el equipo investigador debe tener disponible el Documento de Consentimiento Informado (Hojas de Información y Hojas de Firma) tanto en *galego* como en castellano en el momento de comenzar el reclutamiento.





SERVIZO
GALEGO
de SAÚDE

Xerencia do Servizo
Galego de Saúde



Y HACE CONSTAR QUE:

1.- El Comité Territorial de Ética de la Investigación de Santiago-Lugo cumple tanto en su composición como en sus PNTs los requisitos legales vigentes.

2.- La composición actual del Comité Territorial de Ética de la Investigación de Santiago-Lugo es:

Presidenta

Pilar Rodríguez Ledo. Médico especialista en Medicina Familiar y Comunitaria.

Vicepresidenta

María Mercedes Rodicio García. Médico especialista en Pediatría.

Secretaria

Ana Estany Gestal. Licenciada en Farmacia.

Vicesecretaria

Catalina Caamaño Isorna. Farmacéutica de Atención Primaria.

Vocales

Francisco Campos Pérez. Licenciado en Biología.

Jesús Fernández Álvarez. Miembro lego.

Ricardo García Martínez. Licenciado en Derecho.

Jaime Gulín Dávila. Farmacéutico especialista en Farmacia Hospitalaria.

Jesús Prego Domínguez. Enfermero.

Lorenzo Armenteros del Olmo. Médico especialista en Medicina Familiar y Comunitaria.

Eva Marcos Doldán. Analista-programadora.

Carlos Rodríguez Moreno. Médico especialista en Farmacología Clínica.

Juan Manuel Vázquez Lago. Médico especialista en Medicina Preventiva y Salud Pública.

Para que conste donde proceda, y a petición de quien proceda, en Santiago de Compostela,

La Secretaria del Comité Territorial de Ética de la Investigación de Santiago Lugo,

Documento asinado dixitalmente por:
Ana Estany Gestal (27/01/2021 14:06)
<https://sede.xunta.gal/ceve/index.html#SAOC-C4C4-8C4H-OF40-TG3H-L7H-L-15M-W161-1752-8080-09>



Xacobeo 2021

ANNEXE V: MATERIAL TRANSFER AGREEMENT (MTA) FOR THE USE OF THE hiPSC LINE CTRL. BS PBIPS37-SV4F-1.



MATERIAL TRANSFER AGREEMENT FOR TRANSFER OF MATERIALS TO RECIPIENT SCIENTISTS

In response to FUNDACIÓN PÚBLICA GALEGA DE MEDICINA XENÓMICA (RECIPIENT) acting through Angel Carracedo Álvarez (Recipient Scientist) request for MATERIAL (induced pluripotent stem cells (Ctrl. BS PBIPS37-SV4F-1) for a research program entitled "*megaT: Enfoque integral y traslacional del diagnóstico y tratamiento de trastornos del neurodesarrollo (TEA, TDAH, DI) y del TOC: Una aproximación genética y funcional*", Banco Nacional de Líneas Celulares of Instituto de Salud Carlos III (PROVIDER) and the RECIPIENT agree to the following before the RECIPIENT SCIENTISTS receives the MATERIAL:

1. The above MATERIAL is the property of the PROVIDER and is made available as a service to the research community. Ownership of the MATERIAL shall remain with PROVIDER and transfer of the MATERIAL to the RECIPIENT shall not affect PROVIDER's ownership of the MATERIAL.
2. This MATERIAL is not to be used for diagnostic or therapeutic purposes.
3. The MATERIAL will be used for teaching or non-commercial research purposes. As used herein, non-commercial research purposes specifically excluded sponsored research wherein the sponsor receives a right whether actual or contingent to the results of the sponsored research. The MATERIAL may not be used for commercial purposes or the direct benefit of research sponsor, except as such research sponsor is permitted to use MATERIAL under a separate written agreement with PROVIDER. Specifically MATERIAL shall not be used in a research program where rights (either actual or contingent) have already been granted to a research sponsor who does not have a separate written agreement with PROVIDER permitting such use of MATERIAL.
4. Nothing contained herein shall be considered to be the grant of a commercial license. Furthermore, nothing contained herein shall be construed to be a waiver of PROVIDER's patent rights or PROVIDER's property rights in the MATERIALS.
5. The MATERIAL will not be further distributed to others without the PROVIDER's written consent. The RECIPIENT shall refer any request for the MATERIAL to the PROVIDER. To the extent supplies are

available, the PROVIDER or the PROVIDER scientist agree to make the MATERIAL available, under a separate simple letter agreement to other scientists for teaching or non-commercial research purposes only.

6. The RECIPIENT agrees to acknowledge the source of the MATERIAL in any publications reporting use of it.
7. Upon PROVIDER's written request, RECIPIENT agrees to provide without cost reasonable quantities of any RECIPIENT MATERIALs that it develops from the use of MATERIAL to PROVIDER for non-commercial research purposes only at PROVIDER after RECIPIENT has publicly disclosed or reasonably characterized such RECIPIENT MATERIALs. Recipient also agrees to grant subject to the RECIPIENT's patent rights a non-commercial research license to PROVIDER.
8. Any MATERIAL delivered pursuant to this Agreement is understood to be experimental in nature and may have hazardous properties. THE PROVIDER MAKES NO REPRESENTATIONS AND EXTENDS NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, OR THAT THE USE OF THE MATERIAL WILL NOT INFRINGE ANY PATENT, COPYRIGHT, TRADEMARK, OR OTHER PROPRIETARY RIGHTS. Unless prohibited by law, RECIPIENT assumes all liability for claims for damages which may arise from the use, storage, handling or disposal of MATERIAL except that, to the extent permitted by law, PROVIDER shall be liable to the RECIPIENT when the damage is caused by the gross negligence or willful misconduct of the PROVIDER.
9. The RECIPIENT agrees to use the MATERIAL in compliance with all applicable statutes, regulations and guidelines relating to their handling, use or disposal.
10. The RECIPIENT agrees to use the MATERIAL only for the purposes stated in the research protocol annexed to the request. Specifically, RECIPIENT agrees that its research program will exclude: (i) the mixing of MATERIAL with an intact embryo, either human or non-human; (ii) implanting MATERIAL or products of MATERIAL in a uterus, and (iii) attempting to make whole embryos with MATERIAL by any method. Recipient shall supply an annual certification statement confirming compliance with the restrictions on the use of MATERIAL supplied by PROVIDER. Recipient agrees that MATERIAL is to be returned to PROVIDER or destroyed upon a MATERIAL breach of the terms of this Agreement by RECIPIENT.
11. The MATERIAL is provided with a transmittal fee solely to reimburse the PROVIDER for its preparation and distribution costs. The amount of the fee for this transfer if MATERIAL will be five hundred euros (500 €) per vial, excluding of VAT and the cost of shipping.
12. The RECIPIENT will send to the PROVIDER a copy of any paper published using the MATERIAL.

13.This Agreement shall be construed and governed by the laws of Spain.

14.Given that the available lines may have been deposited by third parties, the PROVIDER does not assume any liability or responsibility in case the properties of any MATERIAL provided does not correspond to the published specifications.

The PROVIDER, RECIPIENT, and RECIPIENT scientist must sign both copies of this letter and return one signed copy to the PROVIDER. The PROVIDER will then send the MATERIAL.

PROVIDER

Name of Authorized Official: M^a Pilar Gayoso Diz
 Titled of Authorized Official: Directora S. G. de Investigación en Terapia Celular y Medicina Regenerativa
 Banco Nacional de Líneas Celulares
 Address: Banco Nacional de Líneas Celulares
 Instituto de Salud Carlos III
 Avda. Monforte de Lemos 5. Pabellón 5. 28029 Madrid. Spain

Signature of Authorized Official.

Date:

RECIPIENT INFORMATION and AUTHORIZED SIGNATURE

Recipient: FUNDACIÓN PÚBLICA GALEGA DE MEDICINA XENÓMICA
 Address: CHUS, Choupana s/n 15706 Santiago de Compostela (A Coruña)
 Name of Authorized Official: Angel Carracedo Alvarez
 Title of Authorized Official: Director Ejecutivo

Signature of Authorized Official.

Date:

Read and understood by:
 Recipient Scientist: Angel Carracedo

CARRACEDO
 ALVAREZ
 ANGEL MARIA
 - 33223018R

Firmado digitalmente
 por CARRACEDO
 ALVAREZ ANGEL
 MARIA - 33223018R
 Fecha: 2022.11.21
 18:40:51 +01'00'

Signature of Recipient Scientist.

Date:

ANNEXE VI: SCIENTIFIC PUBLICATIONS INCLUDED IN THE THESIS.

Title: Experimental Models to Study Autism Spectrum Disorders: hiPSCs, Rodents and Zebrafish.

Authors: Alba Pensado-López^{1,2}, Sara Veiga-Rúa^{1,2}, Ángel Carracedo^{2,3}, Catarina Allegue², Laura Sánchez¹.

Affiliations:

1. Department of Zoology, Genetics and Physical Anthropology, Universidade de Santiago de Compostela, Campus de Lugo, 27002 Lugo, Spain.
2. Genomic Medicine Group, Center for Research in Molecular Medicine and Chronic Diseases (CiMUS), Universidade de Santiago de Compostela, 15706 Santiago de Compostela, Spain.
3. Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), CIMUS, Universidade de Santiago de Compostela, 15706 Santiago de Compostela, Spain.

Publication date: 20/11/2020

Editorial: Multidisciplinary Digital Publishing Institute (MDPI)

DOI: 10.3390/genes11111376.

ISSN: 2073-4425 (Genes)

Impact factor: 3.5

Quartile: Q2

Authorization from the journal/editor for the use of the publication in the doctoral thesis: the article has been published in open access and distributed under a Creative Commons CC BY license. This license allows the use and reproduction of segments of the article without a specific authorization, as long as it is properly cited.

Contribution of the authors to the publication:

1. Introduction.
 - 1.1. Definition and Epidemiology of Autism Spectrum Disorders.
 - 1.2. Aetiology of Autism Spectrum Disorders.
 - 1.3. Diagnostic of Autism Spectrum Disorders.
 - 1.4. Treatment of Autism Spectrum Disorders.

This section was written by Sara Veiga Rúa and Alba Pensado López. and reviewed by the rest of the authors. Figures 1 and 2, and Tables 1 and 2 were created by Sara Veiga Rúa.

2. Genome Editing Systems, a Promising Tool for Modelling Human Disorders.

This section, including Figure 3, was written entirely by Sara Veiga Rúa and reviewed by the rest of the authors.

3. *In Vitro* Models of ASD: The Stem Cell Revolution.

This section, including Table 3, was written entirely by Sara Veiga Rúa and reviewed by the rest of the authors.

4. Animal Models in ASD Research.

4.1. Rodents and the Modelling of Human Disorders.

4.1.1. *Mus Musculus* in ASD Research.

4.1.2. *Rattus norvegicus* in ASD Research.

4.2. Zebrafish and the Modelling of Human Disorders.

4.2.1. Zebrafish and Mammals: Conservation throughout Evolution.

4.2.2. Gene Targeting in Zebrafish.

4.2.3. Characterization of Zebrafish Models.

4.2.4. Limitations of Zebrafish to Model Human Disorders.

This section was written by Sara Veiga Rúa (4, 4.1, 4.1.1 and 4.1.2) and Alba Pensado López (4.2, 4.2.1, 4.2.2, 4.2.3, and 4.2.4) and reviewed by the rest of the authors. Figure 4 and Tables 4 and 6 were elaborated by Sara Veiga Rúa. Tables 5, 7, and 8 were created by Alba Pensado López.

5. Future Challenges.

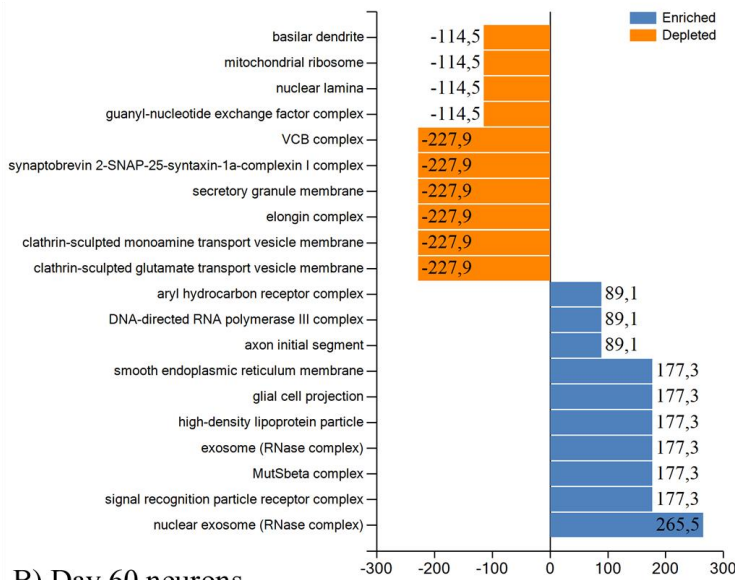
5.1. *In Vitro* Modelling.

5.2. *In Vitro* Modelling

This section was written by Sara Veiga Rúa and reviewed by the rest of the authors.

ANNEXE VII: SUPPLEMENTARY PROTEOMIC DATA.

A) Day 40 neurons



B) Day 60 neurons

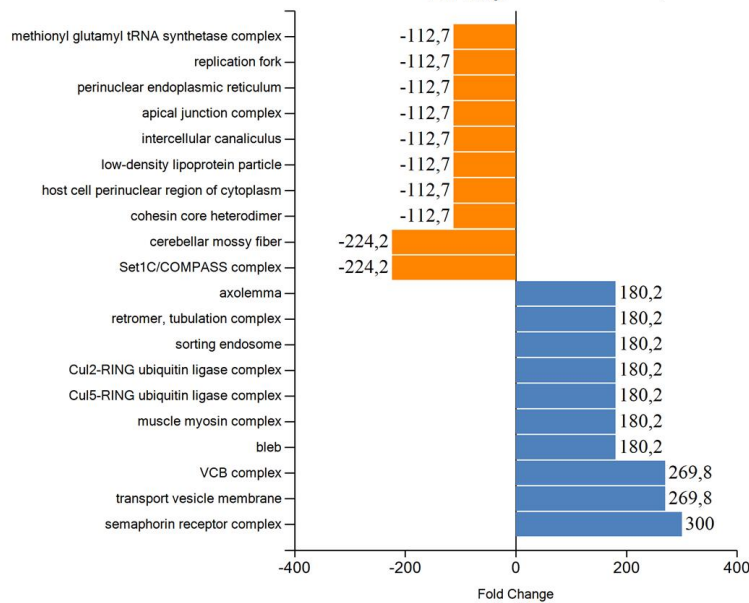


Figure 141. Cellular components enriched (orange) and depleted (blue) in cortical neurons derived from the cell line Ctrl. BS PBiPS37-Sv4F-1-*BTBD8-i2* in comparison to the wild type isogenic control. Comparison between neuronal cultures from day 40 (A) and day 60 (B). Represented the top 10 results obtained with FunRich v3.1.3.

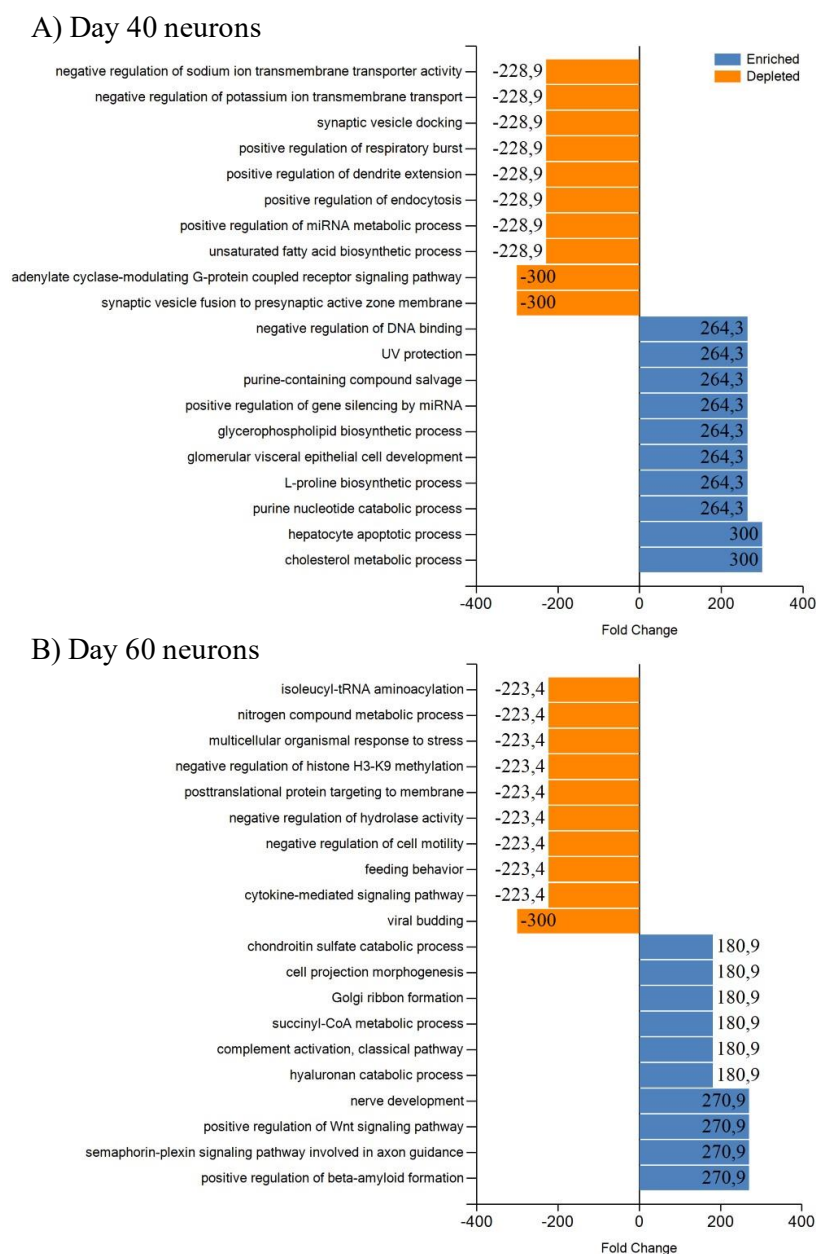
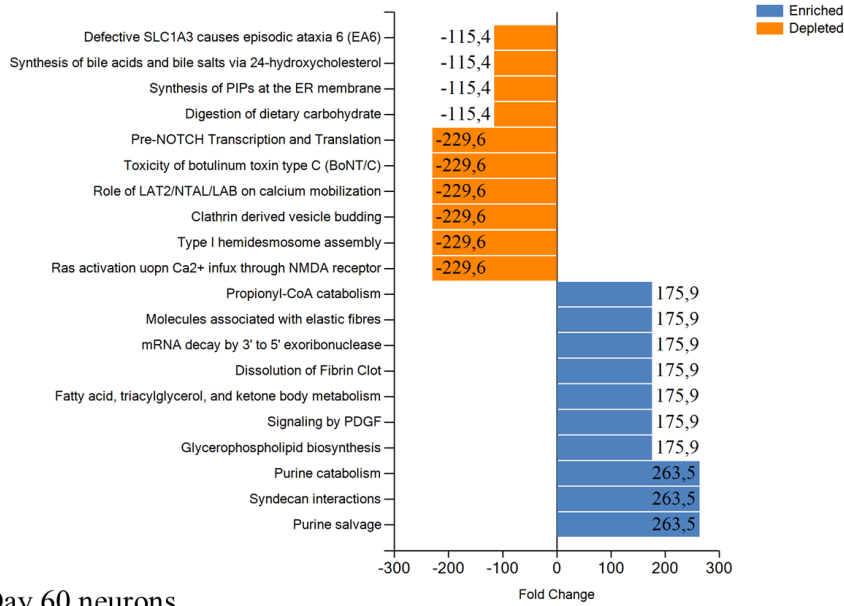


Figure 142. Biological processes enriched (orange) and depleted (blue) in cortical neurons derived from the cell line Ctrl. BS PBiPS37-Sv4F-1-BTBD8-i2 in comparison to the wild type isogenic control. Comparison between neuronal cultures from day 40 (A) and day 60 (B). Represented the top 10 results obtained with FunRich v3.1.3.

A) Day 40 neurons



B) Day 60 neurons

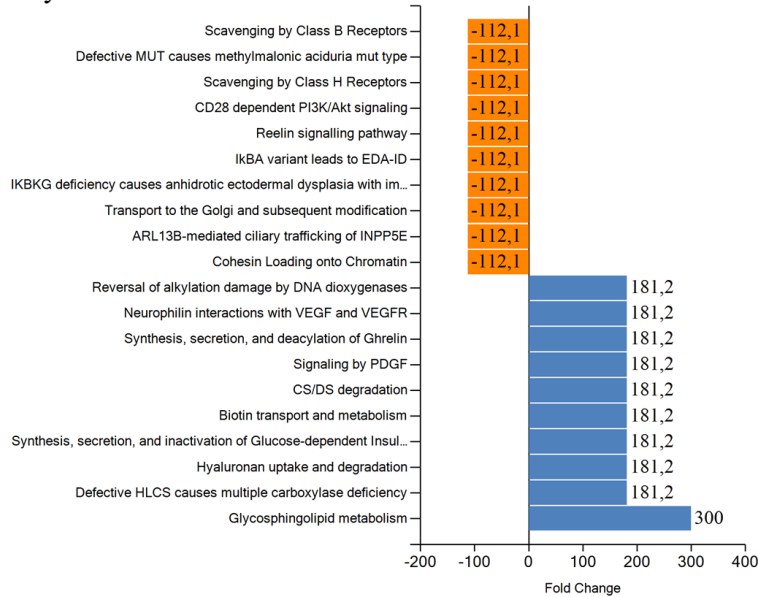
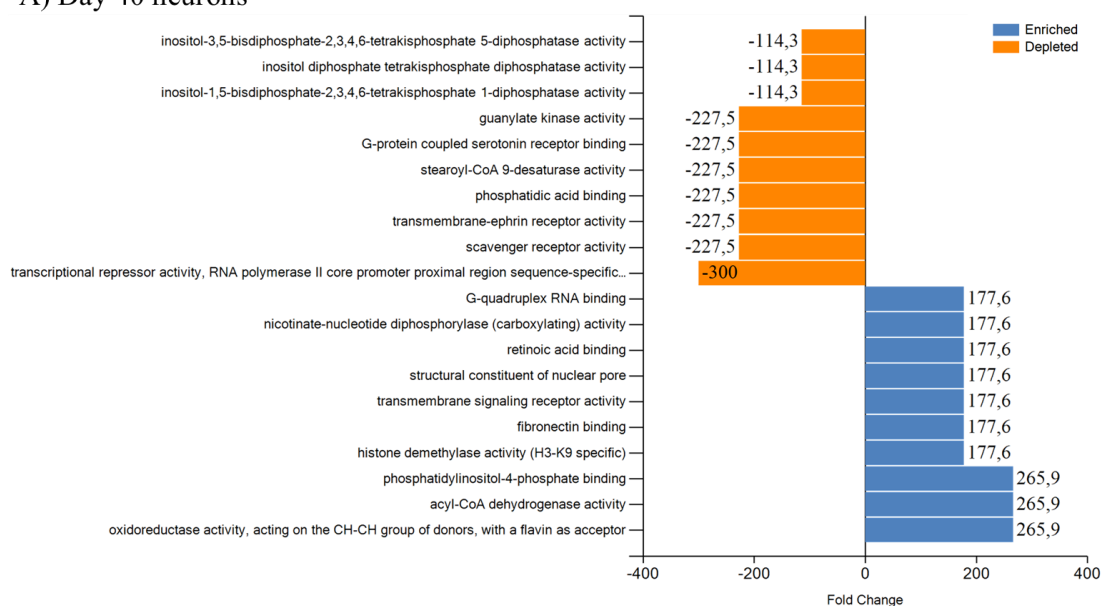


Figure 143. Reactome pathways enriched (orange) and depleted (blue) in cortical neurons derived from the cell line Ctrl. BS PBiPS37-Sv4F-1-*BTBD8-i2* in comparison to the wild type isogenic control. Comparison between neuronal cultures from day 40 (A) and day 60 (B). Represented the top 10 results obtained with FunRich v3.1.3.

A) Day 40 neurons



B) Day 60 neurons

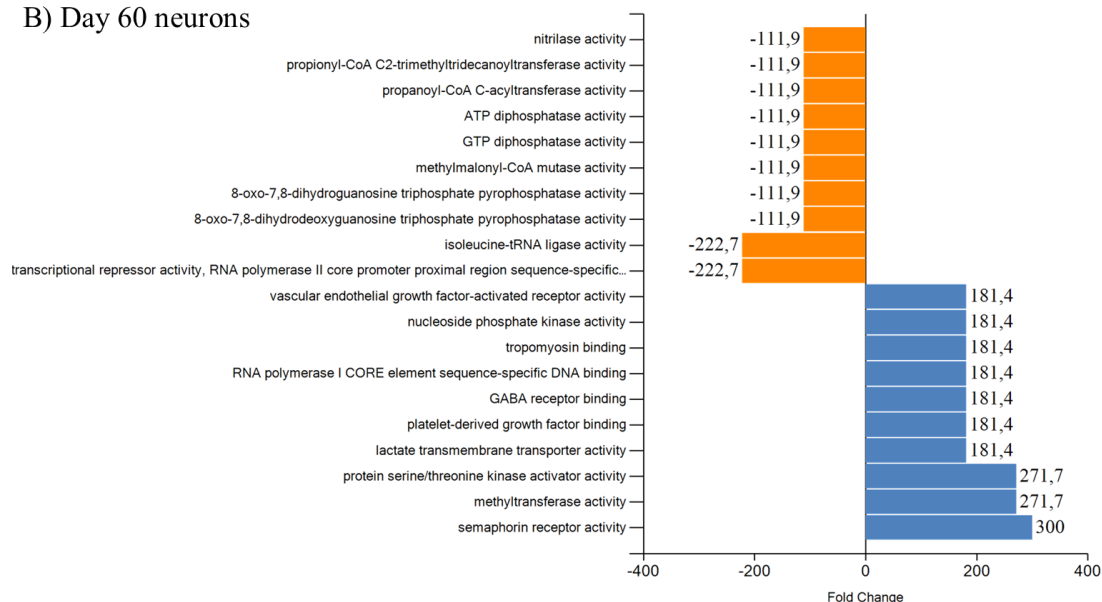


Figure 144. Molecular functions enriched (orange) and depleted (blue) in cortical neurons derived from the cell line Ctrl. BS PBiPS37-Sv4F-1-BTBD8-i2 in comparison to the wild type isogenic control. Comparison between neuronal cultures from day 40 (A) and day 60 (B). Represented the top 10 results obtained with FunRich v3.1.3.



Autism Spectrum Disorders are characterized by behavioural, social, and communication impairments. Despite their prevalence, the precise mechanisms underlying ASD are poorly understood. This thesis is focused on developing suitable models to study the impact of mutations in three genes (BTBD8, FOSL2, and TAOK1) potentially associated with ASD. In vitro models were generated using human iPSCs that were differentiated into cortical neurons and human brain organoids, that were characterized through proteomics and scRNA-seq. Genes of interest were also studied in *Danio rerio*, to evaluate the morphological and behavioural consequences of gene dysfunction. These models will be useful for improving our understanding of the molecular mechanisms underlying ASD, but also for improving the diagnosis and treatment of the disorder.