

Dynamic X-chromosomal reactivation during neuronal differentiation as a mechanism of female resilience

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften

Am Fachbereich Biologie
der Johannes Gutenberg-Universität Mainz

Marco Bertin
geb. am 22.09.1992 in Rom, Italien

Mainz, den

Dekan: Prof. Dr. Eckhard Thines

First supervisor:

Second supervisor:

Date of oral examination:

Table of Contents

Abstract.....	9
Zusammenfassung.....	10
List of abbreviations	11
1 Introduction.....	12
1.1 Induced pluripotent stem cells: Modelling human-specific diseases and developmental processes.....	12
1.2 In vitro modelling of human development.....	13
1.2.1 Early stages of human development	13
1.2.2 Human pluripotent stem cells as a model for early embryonic differentiation.....	15
1.2.3 Naïve versus primed PSCs.....	16
1.2.4 Brain organogenesis and Neurodevelopment.....	17
1.2.5 Early stages of brain development.....	17
1.2.6 Neurogenesis and gliogenesis.....	20
1.2.6.1 Neurogenesis.....	20
1.2.6.2 Gliogenesis.....	21
1.2.7 <i>In vitro</i> modeling of human neuronal differentiation and brain organogenesis.....	21
1.3 X-chromosome Inactivation.....	22
1.3.1 Sexual dimorphism in mammals.....	23
1.3.2 X-chromosome Inactivation as a mechanism of dosage compensation in mammals.....	23
1.3.3 lncRNA <i>XIST</i> expression during early development (initiation of XCI)	24
1.3.4 Initiation of XCI in humans: the dual role of <i>XIST</i> and <i>XACT</i>	26
1.3.5 Role of inactive X-chromosome in sexual dimorphism: escapee genes	28
1.3.6 Methods for investigating escape genes.....	28
1.3.7 Role of escape genes in disease phenotype.....	29
1.4 Neurodevelopmental disorders	30
1.4.1 Sex differences during neurodevelopmental process	30
1.4.2 Sex bias in neurodevelopmental disorders.....	31
1.4.3 Opitz/BBB G Syndrome	31
1.5 MID1	32
1.5.1 Protein function.....	32
1.5.2 MID1 regulates the mTOR pathway	33
1.5.3 Mutations affecting <i>MID1</i> gene	35
2 Aim of the thesis.....	37

3	Material and Methods	38
3.1	Material.....	38
3.1.1	Equipment.....	38
3.1.2	Disposable material.....	39
3.1.3	Chemicals and Media.....	41
3.1.3.1	Chemicals.....	41
3.1.3.2	Media.....	42
3.1.3.3	Media composition.....	42
3.1.4	Reagents, Kits, and Enzymes.....	43
3.1.4.1	Kits and reagents.....	43
3.1.4.2	Enzymes.....	44
3.1.5	Primers and Antibody	44
3.1.5.1	Primers.....	44
3.1.5.2	Primary antibodies	46
3.1.5.3	Secondary antibodies	46
3.1.5.4	RNA-FISH probes	46
3.1.6	Plasmids and gRNA constructs.....	47
3.1.7	Softwares and online tools.....	47
3.1.8	Cells.....	48
3.2	Methods	50
3.2.1	Cell culture.....	50
3.2.1.1	Coatings	50
3.2.1.2	Gelatine.....	50
3.2.1.3	Matrigel.....	50
3.2.1.4	Geltrex	50
3.2.1.5	Poly-Ornithine/Laminin	50
3.2.2	Cell culturing	51
3.2.3	Fibroblasts.....	51
3.2.3.1	Isolation	51
3.2.3.2	Thawing	51
3.2.3.3	Culturing.....	51
3.2.3.4	Splitting.....	51
3.2.3.5	Freezing	52
3.2.4	Induced Pluripotent Stem Cells (iPSCs)	52
3.2.4.1	Thawing	52

3.2.4.2	Culturing and cleaning of spontaneously differentiated cells	52
3.2.4.3	Splitting.....	52
3.2.4.4	Colony picking.....	53
3.2.4.5	Freezing	53
3.2.4.6	Genome editing.....	53
3.2.5	Neural progenitor and stem cells (NPCs).....	53
3.2.5.1	Differentiation of iPSCs into NPCs	53
3.2.5.2	Culturing	54
3.2.5.3	Splitting.....	54
3.2.5.4	Freezing	54
3.2.5.5	Thawing	55
3.2.6	Neurons.....	55
3.2.6.1	Differentiation and culturing.....	55
3.2.7	Cell culture methods	55
3.2.7.1	Cell pellet collection	55
3.2.7.2	PFA fixation of adherent cells	55
3.2.7.3	Cell counting.....	56
3.2.7.4	Karyotyping	56
3.2.8	Molecular methods	56
3.2.8.1	Immunofluorescence staining	56
3.2.8.2	RNA-FISH procedure and imaging	56
3.2.8.3	Quantification of RNA-FISH signals.....	57
3.2.8.4	DNA extraction.....	57
3.2.8.5	Total RNA isolation.....	58
3.2.8.6	Retro-transcription of total RNA and cDNA synthesis.....	58
3.2.8.7	Sendai Virus vector clearance.....	59
3.2.8.8	Allele-specific RT-PCR.....	59
3.2.8.9	Quantification of Allele-Specific Expression by Pyrosequencing (QUASEP)	60
3.2.8.10	RT-qPCR	61
3.2.9	RNA-sequencing.....	62
3.2.9.1	RNA-sequencing library preparation	62
3.2.9.2	RNA-sequencing data pre-processing.....	63
3.2.9.3	Allele-specific expression analysis (ASE) of the bulk RNA-sequencing data	63
3.2.9.4	Estimation of biallelic expression of X-linked genes	64
3.2.9.5	Generation of X-chromosome ideograms	64

3.2.9.6	Overlap analysis with the known escape genes in humans and mouse	64
3.2.9.7	Enrichment analysis of Neurodevelopmental Disorder-associated genes	65
3.2.9.8	Cluster analysis of biallelically expressed genes.....	65
3.2.9.9	Protein-protein interaction network analysis of the reactivated genes	65
3.2.9.10	Chromatin state sex-specific analysis of the reactivated and silenced genes	66
3.2.9.11	Sex-biased tissue-specific expression of reactivated, late-silenced and full-escapee genes	66
3.2.9.12	Disease-ontology enrichment analysis.....	66
4	Results	67
4.1	Characterization of cell lines used for the study	67
4.1.1	Induced pluripotent stem cells: confirmation of pluripotency identity and clonal status ..	67
4.1.2	Neuronal differentiation of iPSCs.....	70
4.1.2.1	Neural progenitor cells (NPCs) differentiation	70
4.1.2.2	Neuronal differentiation of NPCs	72
4.2	2. Escape gene expression during neuronal differentiation.....	73
4.2.1	Dynamic expression from the inactive X-chromosome.....	74
4.2.2	Validation of predicted reactivated genes	82
4.2.2.1	QUASEP assay results confirmed reactivation expression pattern of the <i>GPM6B</i> gene.....	82
4.2.2.2	Allele-specific RT-PCR and QUASEP assay confirmed <i>MIDI</i> reactivation.....	85
4.3	Erosion of the X-chromosome is not responsible for the dynamic gene expression from the inactive X-chromosome	91
4.3.1	RNA-seq data revealed heterogeneous presence of eroded and non-eroded cell lines ..	91
4.3.2	Cell line screening for hallmark of erosion revealed complex pattern of erosion in iPSCs and NPCs.....	92
4.3.3	Single-nuclei validation of dynamic expression from the inactive X-chromosome by using RNA-FISH	98
5	Discussion.....	106
5.1	In vitro differentiation of human pluripotent stem cells can faithfully recapitulate the human neurodevelopmental process.....	106
5.1.1	Reprogramming of somatic cells into iPSCs.....	107
5.1.2	iPSC differentiation into NPCs	108
5.1.3	NPC differentiation into neurons	109
5.2	Usage of the in-vitro system to study facultative escape from X-inactivation.....	110
5.2.1	Skewed in vitro system can reliably identify facultative escape genes	110
5.2.2	Erosion of the inactive X-chromosome fails to explain the dynamic expression of facultative escape genes	111

5.3	Identification of novel categories of facultative escapees	115
5.3.1	Technologies for the identification of escape genes	115
5.3.2	Constitutive and facultative escape gene expression across <i>in vitro</i> neural differentiation	117
5.3.3	RNA-FISH nuclear signal distribution recapitulates the dynamic X-chromosomal gene expression	118
5.3.4	Cell type specific reactivation of genes from the inactive X-Chromosome	120
5.3.5	Xp22.2 region: a hotspot for reactivated genes	122
5.3.6	Time and cell lineage specific reactivation of <i>MIDI</i> influences the phenotype of OS syndrome	123
6	Conclusion	128
7	References	129
8	Curriculum vitae	155
9	Acknowledgements	158

Abstract

In humans, females possess two X-chromosomes, while males have one X and one Y chromosome, necessitating a mechanism to balance the expression of X-linked genes. This is achieved through X-chromosome inactivation (XCI), a process involving chromatin and epigenetic modifications, as well as nuclear re-localization, which leads to the transcriptional silencing of one randomly selected X chromosome. XCI is established early in human epiblast cells after the implantation stage of the embryo and is maintained throughout the lifespan of female individuals, with the inactive X-chromosome inherited by daughter cells.

However, XCI is incomplete, with approximately 20% of X-linked genes being either partially or fully expressed from the inactive X chromosome. These genes include constitutive escape genes, which are consistently expressed, and facultative escape genes, which are variably inactivated in a tissue-specific manner. Neurodevelopmental disorders (NDDs) exhibit a notable sex bias, with males more frequently and severely affected than females, although the underlying mechanisms remain unclear. The presence of an additional X chromosome in females and the enrichment of NDD-associated genes on the X chromosome may partly explain these sex differences, with tissue-specific escape genes likely playing a key role.

In this study, we leveraged an induced pluripotent stem cell (iPSC) in vitro system to model post-implantation epiblast cells, characterized by stable XCI. We confirmed the clonal status of the iPSCs, with the same X chromosome either consistently active or inactive across the population, and established an in vitro differentiation protocol to mimic human neurodevelopment. The iPSCs were differentiated into neural stem and progenitor cells (NPCs) and neurons, and the transcriptome was analyzed at allele-specific resolution by tracking the expression of heterozygous genomic variants.

We identified three categories of escape genes based on their allele-specific expression in iPSCs, NPCs, and neurons. Constitutive escape genes, referred to as full-escape genes, were biallelically expressed across all three cell types. We also discovered two novel categories of facultative escape genes: (1) reactivated genes, which escape XCI specifically in neuronal cells, and (2) late-silenced genes that escape XCI in iPSCs but become inactivated during neuronal differentiation. Follow-up analyses of neuronal-specific reactivated genes support the idea that these genes may play a crucial role in explaining the female protection against NDDs. Additionally, we used RNA-FISH technology to investigate the reactivation mechanism at the single-cell level, confirming the neuronal-specific reactivation of two candidate genes, *MIDI1* and *GPM6B*.

In summary, the data collected during my PhD demonstrate dynamic expression patterns from the inactive X chromosome, with certain genes being reactivated or silenced during neuronal differentiation. These findings suggest that differentiation-induced use of the inactive X chromosome may contribute to the observed sex-biased phenotypes in NDDs between males and females.

Zusammenfassung

Beim Menschen besitzen Frauen zwei X-Chromosomen, während Männer ein X- und ein Y-Chromosom aufweisen, was einen Mechanismus zur Ausbalancierung der Expression von X-chromosomalen Genen erfordert. Dies wird durch die X-Chromosomen-Inaktivierung (XCI) erreicht, einem Prozess, der chromatin- und epigenetische Modifikationen sowie die nukleare Relokalisierung umfasst und zur transkriptionellen Inaktivierung eines zufällig ausgewählten X-Chromosoms führt. Die XCI wird früh in menschlichen Epiblastzellen nach der Implantationsphase des Embryos festgelegt und während des gesamten Lebens der weiblichen Individuen aufrechterhalten, wobei das inaktive X-Chromosom an Tochterzellen weitergegeben wird.

Die XCI ist jedoch unvollständig, da etwa 20 % der X-chromosomalen Gene teilweise oder vollständig vom inaktiven X-Chromosom exprimiert werden. Zu diesen Genen gehören konstitutive Escape-Gene, die konstant exprimiert werden, und fakultative Escape-Gene, die in einem gewebespezifischen Muster inaktiviert werden. Neuroentwicklungsstörungen (NDDs) zeigen einen deutlichen Geschlechtsunterschied, bei dem Männer häufiger und schwerer betroffen sind als Frauen, obwohl die zugrunde liegenden Mechanismen noch unklar sind. Das zusätzliche X-Chromosom bei Frauen und die Anreicherung von NDD-assoziierten Genen auf dem X-Chromosom könnten diese Geschlechtsunterschiede teilweise erklären, wobei gewebespezifische Escape-Gene eine Schlüsselrolle spielen könnten.

In dieser Studie nutzten wir ein in vitro-System mit induzierten pluripotenten Stammzellen (iPSCs), um menschliche Epiblastzellen nach der Implantation zu modellieren, die durch stabile XCI gekennzeichnet sind. Wir bestätigten den klonalen Status der iPSCs, wobei dasselbe X-Chromosom in der Zellpopulation entweder konstant aktiv oder inaktiv war, und etablierten ein Differenzierungsprotokoll in vitro, um die menschliche Neuroentwicklung nachzuahmen. Die iPSCs wurden in neuronale Stamm- und Vorläuferzellen (NPCs) sowie Neuronen differenziert, und das Transkriptom wurde auf allelspezifischer Ebene durch die Verfolgung der Expression heterozygoter Genvarianten analysiert.

Wir identifizierten drei Kategorien von Escape-Genen basierend auf ihrer allelspezifischen Expression in iPSCs, NPCs und Neuronen. Konstitutive Escape-Gene, sogenannte vollständige Escape-Gene, wurden biallelisch in allen drei Zelltypen exprimiert. Darüber hinaus entdeckten wir zwei neuartige Kategorien von fakultativen Escape-Genen: (1) reaktivierte Gene, die speziell in Neuronen der XCI entkommen, und (2) spät-stillgelegte Gene, die in iPSCs der XCI entkommen, aber während der neuronalen Differenzierung inaktiviert werden. Weiterführende Analysen neuronenspezifisch reaktiver Gene unterstützen die Idee, dass diese Gene eine wichtige Rolle bei der Erklärung des weiblichen Schutzes vor NDDs spielen könnten. Zusätzlich setzten wir RNA-FISH-Technologie ein, um den Reaktivierungsmechanismus auf Einzelzellebene zu untersuchen und bestätigten die neuronenspezifische Reaktivierung von zwei Kandidatengenen, MID1 und GPM6B.

Zusammenfassend zeigen die während meiner Promotion gesammelten Daten dynamische Expressionsmuster vom inaktiven X-Chromosom, wobei bestimmte Gene während der neuronalen Differenzierung reaktiviert oder stillgelegt werden. Diese Ergebnisse deuten darauf hin, dass die differenzierungsinduzierte Nutzung des inaktiven X-Chromosoms zu den geschlechtsabhängigen Phänotypen von NDDs bei Männern und Frauen beitragen könnte.

List of abbreviations

iPSCs – induced pluripotent stem cells	UL – upper layer
ESCs – embryonic stem cells	OPCs - oligodendrocyte progenitor cells
NPC – neural progenitor cells	CNS – central nervous system
CRISPR - clustered regularly interspaced short palindromic repeats	ECMs - extracellular matrices
ICM - inner cell mass	SMAD – small mother against decapentaplegic
scRNA-seq - Single cell RNA-seq	XCI – X-chromosome inactivation
dpf - days post-fertilization	lncRNA – long non-coding RNA
PS - primitive streak	XCR – X-chromosome reactivation
EMT - epithelial-mesenchymal transition	IVF – <i>in vitro</i> fertilization
hPSC – human pluripotent stem cell	<i>XIST</i> – X-inactive specific transcript
LIF - Leukemia Inhibitor Factor	XCI – X-inactivation centre
EpiSCs - epiblast stem cells	RNA-FISH – fluorescent in situ hybridization
mESCs – mouse embryonic stem cells	<i>XACT</i> – X-active specific transcript
NEC - neuroepithelial cells	Xi – inactive X-chromosome
VZ - ventricular zone	Xa – active X-chromosome
vRG - ventricular radial glial cells	ID – intellectual disabilities
aRG - apical radial glial cells	NDD – neurodevelopmental disorder
SVZ – subventricular zone	ASD – autism spectrum disorder
BPs – basal progenitors	ADHD – attention deficit hyperactivity disorder
IFL - inner fiber layer	SCA – sex chromosome aneuploidy
bIPCs - basal intermediate progenitor cells	FXS – fragile X syndrome
oRG – outer radial glia	XLOS – X-linked Opitz syndrome
PCW – post-conceptual weeks	mTOR – mammalian target of Rapamycin
tRG – truncated radial glial cells	QUASEP – quantification of allele-specific expression by pyrosequencing
NE – neuroepithelial cells	
CR – Cajal-Retzius	
SP – subplate cells	
MZ – marginal zone	
DL – deep layer	

1 Introduction

1.1 Induced pluripotent stem cells: Modelling human-specific diseases and developmental processes

Induced Pluripotent Stem Cells (iPSCs) are an artificial type of pluripotent stem cells not naturally found in the body but obtained from different sources of somatic cells. In 2006, a pioneering article by Yamanaka demonstrated that somatic cells can be reprogrammed by the expression of four transcription factors (referred to as OSKM): Octamer 3/4 (Oct3/4), Krüppel-like factor 4 (Klf4), SRY-box containing gene 2 (Sox2), and the proto-oncogene cytoplasmic Myc protein (c-Myc). Consequently, the terminally differentiated state of the cell can be reverted to a pluripotent stem cell capable of differentiating into any embryonic lineage¹.

After the groundbreaking work by Takahashi and Yamanaka, in which fibroblasts were transduced with retroviruses to convert them into pluripotent stem cells, efforts have been directed towards reprogramming various somatic cell sources, as well as increasing the reprogramming efficiency and safety through the utilization of different reprogramming technologies. To date, iPSCs can be derived from a diverse array of somatic tissues and cell sources. This includes terminally differentiated cells as well as somatic stem cells, senescence cells and cells from aged donors, and even cells obtained from biological waste materials and urine^{2,3}. The advantage of this wide range of cell sources lies in its ability to generate iPSCs in scenarios where patient-specific iPSCs are required but the donor cell source is limited. However, the choice of the donor cells for reprogramming significantly influences the characteristics of the resulting iPSC. Studies have indicated that the efficiency of reprogramming, typically achieving up to 4%, varies depending on the cell type of origin⁴. Furthermore, the reprogramming of somatic cells into iPSCs is not completed: iPSCs retain an epigenetic memory derived from the original cells of origin that in turn influences the ability of the iPSCs to differentiate into any cell type⁵. Among the most common cell sources for reprogramming are blood cells and skin fibroblasts. This preference is due to the availability of highly standardized reprogramming methods and cell culturing conditions, easy access to these cells, and the establishment of biobanks containing iPSC lines derived from both healthy individuals and patients. These biobanks also provide a readily accessible source of blood cells and skin fibroblasts as donor cells for reprogramming.

Nearly two decades following the discovery of iPSCs, extensive literature and established protocols for differentiating them into nearly all cell lineages have illuminated the potential of this technology for modelling human development and diseases⁶⁻¹², as well as the iPS application in medical research¹³. This advancement has led to a shift away from relying solely on model organisms, with iPSCs and their differentiation capabilities offering a means to study the human developmental process more comprehensively¹⁴. By deriving iPSCs from patients, novel models of human diseases have been created, enabling the investigations into both cellular-level and complex disease phenotypes by differentiating iPSCs into specific cell types or 3D organoid systems¹⁵. The derivation of human-specific cell types and even more complex tissue-specific organoids has opened a new era for high-throughput toxicology screening of new compounds and the discovery of novel drug treatments¹⁶⁻¹⁸. Moreover, the reprogramming of patient cells, coupled with novel genetic engineering techniques such as CRISPR-Cas9 technology, has unlocked a promising strategy for autologous cell-based therapy using healthy stem-cell or stem cell-derived differentiated cells¹⁹.

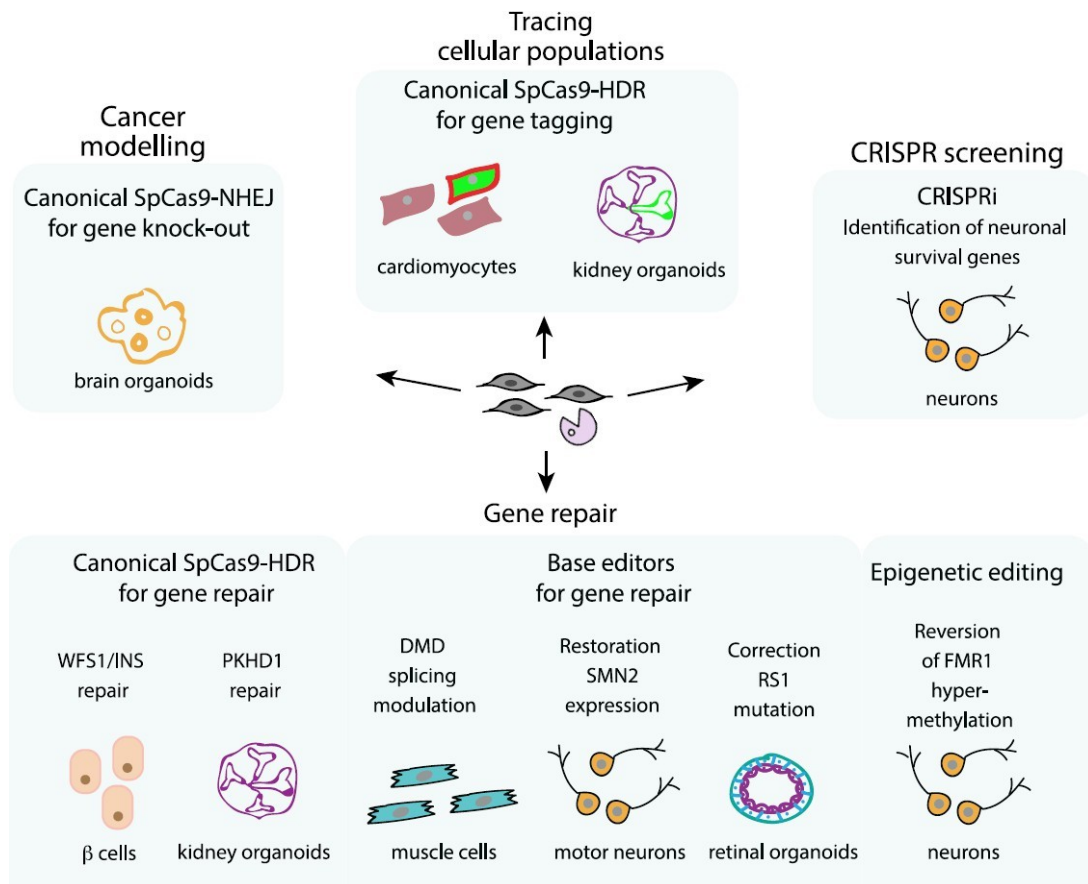


Figure 1. CRISPR-Cas9 strategies applied to hiPS cultures. Figure adapted from Hendriks et al. 2020²⁰. Different approaches of CRISPR technology have facilitated the correction or introduction of patient-derived mutations, enabling precise genome editing. Additionally, CRISPR techniques have been used in tracing specific cell populations or gene expression patterns, as well as genetic screening to identify cell-specific survival genes.

1.2 In vitro modelling of human development

Studies of human embryonic development traditionally relied on model organisms such as amphibians or transgenic mice to understand the fundamental mechanisms underlying embryonic development. However, species-specific differences are fundamentally important in the context of medicine²¹. Recent research has revealed primate and human-specific features during the early stages of development^{22–25}. These findings highlight the importance of comprehending the human-specific developmental processes and emphasize the necessity of improving human-specific models for more accurate insights into human biology and pathology²⁶.

1.2.1 Early stages of human development

The earliest human differentiation stage is called the pre-implantation stage, lasting for 6 days post-fertilization (dpf). It encompasses the development of the zygote (1 dpf), the cell resulting from the fusion of the two gametes, into the blastocyst (5–6 dpf), a sphere-like structure composed of the first two embryonic lineages: trophoblast cells and inner cells within a cavity called the inner cell mass (ICM), which later differentiates into the actual embryo. Comprehensive analysis of early developmental stages has revealed differences between human and mouse development²⁷. At 3.5 dpf, after zygote genome

activation, the morula stage begins, with the embryo consisting of 16 cells. Transcriptomic analyses of the morula have shown early specification of embryonic cells into trophectoderm, expressing *Cdx2* (caudal-type homeobox 2) and *GATA3* (GATA binding protein 3), while ICM cells express *Oct4* (octamer-binding transcription factor 4). During the morula stage, an apico-basal polarization process plays a crucial role in the differentiation fate of embryonic cells: the polarization of apical cells is induced by an increase in the F-actin protein following an accumulation of *PARP* mRNA. Initially, all cells express *GATA3*, but at the end of the polarization process, only apical cells retain *GATA3* expression, becoming trophoblast cells. Apolar cells, which do not undergo polarization, repress *GATA3* expression and form the ICM²⁸. Ultimately, after the polarization process, the morula develops into the blastocyst (5 dpf), characterized by the presence of a cavity as a result of the compaction process. Between 6 and 7 dpf, blastocyst cells are divided into three lineages, characterized by the expression of specific transcription factors: in addition to the trophoblast cells expressing the transcription factors mentioned above, epiblast is characterized by the expression of *NANOG* (derived from the Irish gaelic *Tir na nÓg*, that means “Land of the Young”), while hypoblast or primitive endoderm expresses *GATA4* and *GATA6*²⁹. An interesting difference between mouse and human development lies in the signaling pathway that leads to the segregation of the ICM into epiblast and hypoblast, the FGF/ERK2 signaling pathway: while in mice this pathway guides the segregation process, in humans, the inhibition of this pathway does not alter the process^{30–32}. Single cell RNA-seq (scRNA-seq) studies in humans have identified that during the development of the morula into a blastocyst, there may be an interaction between the epiblast and trophectoderm important for the maturation of the epiblast³³.

Around 6.5 to 7.5 days post-fertilization (dpf), the blastocyst undergoes the implantation process, during which trophoblast cells adhere to and invade the uterine tissue. In humans, during the implantation stage, the trophoblast region responsible for attaching to the endometrium is adjacent to the epiblast, while in mice, the trophoblast region with the same function is opposite to the epiblast³⁴. This suggests a human-specific interaction between the polar trophectoderm and epiblast with a potential role during the implantation stage. Once inside the uterus, the blastocyst enters the peri-implantation stage (6.5-7.5 to 14 dpf). As the blastocyst develops, epiblast cells lose their totipotency and acquire a pluripotent state, meaning they can differentiate into any embryonic lineage (ectoderm, mesoderm, and endoderm), including germ cells but not into extraembryonic tissue³⁵. Studies on pluripotent stem cells in humans and mice have revealed another striking difference between the epiblast pluripotency state in the two species: while mouse epiblast cells cannot differentiate into extraembryonic cells, human pluripotent cells can form extraembryonic tissues under certain conditions^{36,37}. In human, the pluripotency of the epiblast lasts until the gastrulation stage (14 dpf). In mice, the implantation process takes place around 5 dpf, with the next stage, gastrulation, starting at 6.5 dpf, meaning that the pluripotency stage lasts only one day. The prolonged pluripotency period present in humans allows for the development of other cell lineages beyond those mentioned above, such as extraembryonic mesoderm and amniotic epithelium, which are not present and needed in mice³⁸.

After the implantation stage, the embryo undergoes profound changes during the gastrulation stage, with epiblast cells rearranging their morphological and molecular features, leading to the establishment of the three embryonic layers and axial organization. The primitive streak (PS) region arises from the epiblast, and sends cells migrating underneath the epiblast and colonizing both the hypoblast, becoming the endoderm, and the space between the epiblast and endoderm, forming the mesoderm. Cells involved in the PS undergo an epithelial-mesenchymal transition (EMT) for the formation of mesoderm and endoderm tissues. After 14 dpf, the embryo fully develops into a gastrula, from which the name of the gastrulation period originates. Only recently has a study elucidated the full transcriptome of the human gastrula at the single cell level, using a single human XY embryo at day 16-19 dpf and describing the

transcriptomes of the different cell lineages³⁹. However, due to ethical and practical difficulties, molecular and cellular characteristics of human embryos post-implantation remain a mystery. By combining different datasets from several studies, the cellular transcriptome of the human embryo at different stages could be derived^{40,41}; this will allow us to reconstruct developmental pathways along specific differentiation lineages, although the scarcity of human embryo studies and samples makes this possible only through bioinformatic approaches and prediction modelling. To empirically validate these predictions, various approaches should be considered, such as the direct use of Embryonic Stem Cells (ESCs) or induced pluripotent stem cells.

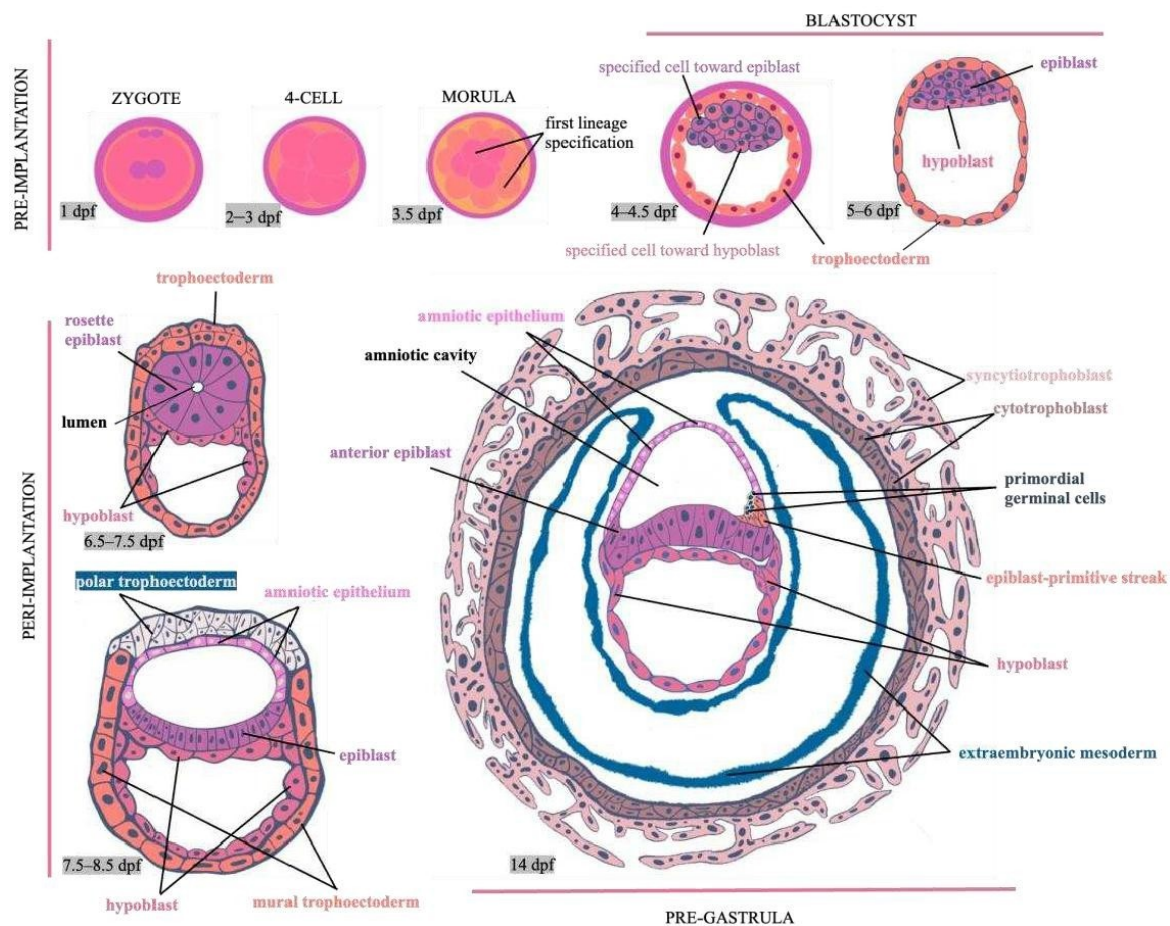


Figure 2. Early stages of human embryonic development. Figure adapted from Ávila-González et al. 2023²⁷. Schematic pictures of the early stages of human embryonic development from the zygote stage after fecundation, to the pre-gastrula stage after implantation.

1.2.2 Human pluripotent stem cells as a model for early embryonic differentiation

Advantages of *in vitro* culturing technologies have made the isolation and expansion of human Embryonic Stem Cells (hESCs) possible, which are self-renewing pluripotent cells derived from the inner cell mass (ICM) or the epiblast of the blastocyst⁴². ESCs were first isolated from mice in 1981^{43,44} (mESCs), and years later from primates^{45,46}. Human ESCs were finally isolated in 1998 from frozen embryos that were generated in excess during reproduction treatments and donated after informed consent from donors⁴⁷. With the discovery of reprogramming technology, induced Pluripotent Stem

Cells (iPSCs) were derived from somatic cells. Under appropriate culture conditions, these cells were shown to have similar characteristics as human ESCs⁴⁸. Since the isolation of human-specific ESCs in 1998, efforts have been made for their characterization and the establishment of differentiation protocols⁴⁹⁻⁵¹, opening the possibility of using hPSCs (hESCs and hiPSCs) for investigating human development. However, concerns arose regarding the isolation method of hES cells⁵², limiting its role in modelling disease and the study of patient-specific mutations. Patient-derived ESCs can be obtained in particular cases after genetic screening of pre-implantation embryos⁵³, or the patient mutation can be introduced by genetic manipulation⁵⁴. To overcome the ethical concerns associated with ESCs isolation and to better study human development with a patient-specific background, iPSCs are widely used for modeling the developmental process replacing the use of hESCs.

1.2.3 Naïve versus primed PSCs

The first attempt to cultivate ESCs was conducted using mouse cells and involved a media formulation containing fetal bovine serum^{55,56}. When maintained in this medium, ES cells exhibited a heterogeneous population, likely due to spontaneous differentiation, and their molecular profile did not match that of embryonic cells. Consequently, this state was termed metastable⁵⁷. Subsequent efforts focused on finding a media composition capable of maintaining cellular characteristics similar to those of the *in vivo* epiblast cells. For mESCs, the addition of Leukemia Inhibitor Factor (LIF), along with other compounds such as PD0325901 and CHIR99021 (ERK2 and GSK3 kinase inhibitors, referred to as LIF/2i condition) helped to sustain the cells in a naïve state, resembling the pluripotency of the pre-implantation epiblast^{58,59}. In human PSCs, the media composition for the naïve state slightly differs and includes formulations like 5i/L/A or t2i/L/Gö⁶⁰. Epiblast cells isolated from mouse embryos after implantation gave rise to epiblast stem cells (EpiSCs). EpiSCs maintain pluripotency, although their molecular signature differs from the naïve state. They were defined as “primed” pluripotency^{61,62}. Differences in gene expression (especially transcription factors), epigenetic signatures, X-chromosome state, and morphology exist between naïve and primed PSCs⁶³⁻⁶⁵; the most significant functional difference is observed during the chimera formation assay, where naïve cells fully re-integrated into a blastocyst, while primed cells did not⁶⁶. Pathways involved in maintaining the undifferentiated state differ between mouse and human: naïve mESCs required the 2i/LIF pathway⁶⁷, while human naivety relies on 5i/L/A or t2i/L/Gö pathways. Additionally, naïve hPSCs express markers of trophoblast cells (TFAP2C, TEAD4, GATA3, and KRT7), and possess the ability to, among others, differentiate into trophoblast-derived specialized cells⁶⁸⁻⁷⁰ (syncytiotrophoblast or extravillous trophoblast). Other pathways, such as Nodal, WNT, and LIF pathways, influence the differentiation of naïve hPSCs into primitive endoderm or hypoblast cells, characterized by specific markers like PDGFRA, GATA6, and NID2⁷¹. Naïve mESCs, however, are unable to differentiate into trophoblast or hypoblast cells. The steamininess of mouse EpiSCs is regulated by the FGF and Activin/Nodal pathways^{72,73}, which also regulate hEpiSCs⁷⁴ (referred to as primed hPSCs). However, hEpiSCs cannot contribute to chimera formation, and transcriptome analysis has revealed gene expression characteristics of human post-implantation embryo at 10-14 dpf, confirming their primed state⁷⁵.

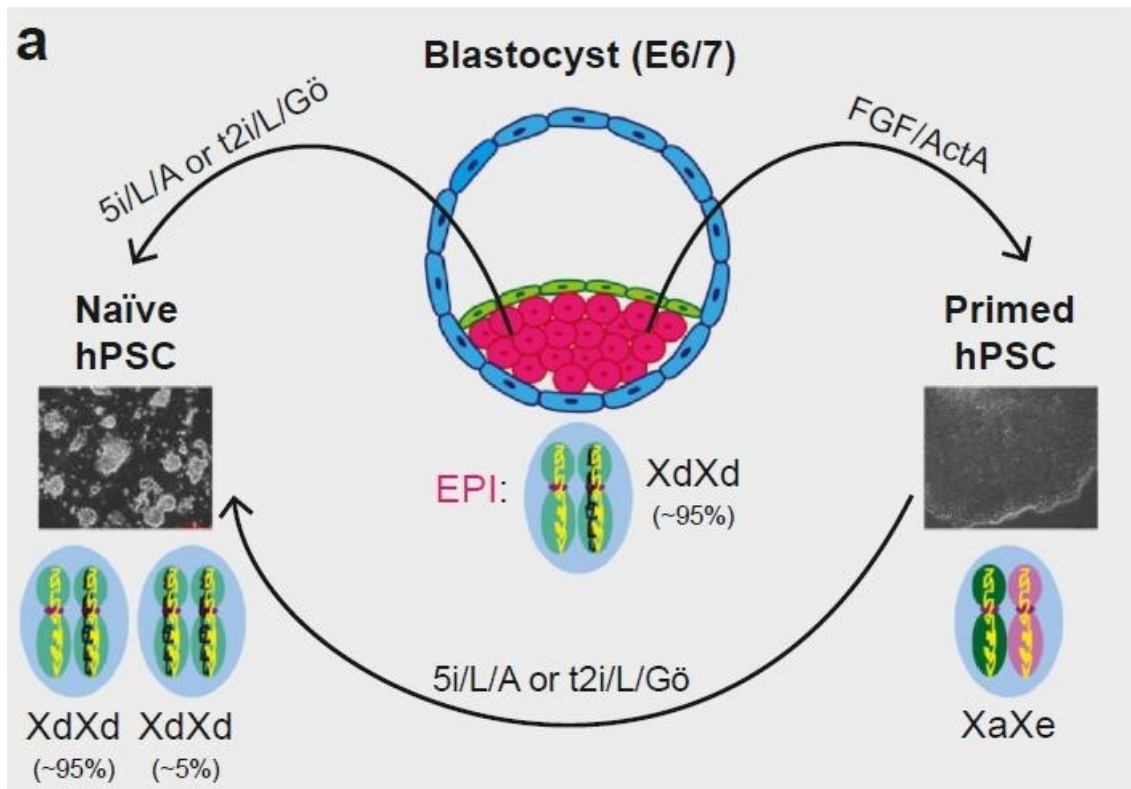


Figure 3. Derivation of naïve and primed hPSCs from different stages of embryonic development. Figure adapted from Khan and Theunissen 2023⁷⁶. Human pluripotent stem cells can be directly derived from human blastocyst or reprogrammed *in vitro*. Media composition change between naïve and primed cells.

1.2.4 Brain organogenesis and Neurodevelopment

Human brain organogenesis is the process that results in the formation of the mature brain from the development of embryonic neuroectoderm. It commences after the gastrulation phase, during which a small number of neuroepithelial cells begin to develop. This process continues throughout pregnancy and even further into childhood, ultimately leading to the formation of the most complex human organ: the brain, which is composed of more than 170 billion cells⁷⁷.

1.2.5 Early stages of brain development

After gastrulation, the embryo is comprised of three layers: ectoderm, mesoderm, and endoderm. Neural tissue development commences during neurulation, where the neural plate derived from the ectoderm folds itself to form the neural tube, the precursor of the entire nervous system, including the brain. The anterior (cephalic) portion of the neural tube consists of a pseudo-stratified monolayer of cells called neuroepithelial cells (NEC) which delimitates a cavity known as the ventricular zone (VZ); NEC primarily undergo symmetric division, expanding their population, which in turn influences the dimension of the brain by creating the cortical neural progenitor pool⁷⁸⁻⁸¹. During organogenesis, as the anterior neural tube thickens due to continued cell divisions, a new cell type emerges from NECs known as ventricular radial glial cells (vRG, apical progenitors, or apical radial glial cells, aRG). These cells have their cell body in the ventricular zone, an apical process in contact with the VZ cavity, and the basal extension traversing the entire cortical thickness^{82,83}. Apical progenitor cells are crucial during early cortical development, as they can undergo different types of cell division: self-renewal symmetric

division or asymmetric division, with the latter giving rise to intermediate progenitor cells (IPCs) or neurons. Symmetric cell division of vRG leads to an increase in progenitor cells and to the formation of the subventricular zone (SVZ) in the mid-gestation period, populated by basal progenitors⁸⁴ (BPs), which mainly drive neuron differentiation⁸⁵⁻⁸⁷. Expansion of the SVZ leads to the formation of the inner and outer SVZ (iSVZ and oSVZ), separated by the inner fiber layer (IFL), axonal fibers forming a thin layer. The oSVZ is a unique feature of mammals with large cortical expansion and contains basal intermediate progenitor cells (bIPCs) and a special type of progenitor cells called outer radial glia (oRG), also known as basal radial glia progenitors, derived from the differentiation of BPs⁸⁸. oRG cells extend their basal processes similarly to vRG cells^{89,90} but are generally not in contact with the VZ. Studies in primates described different forms of those cells⁹⁰⁻⁹³. oRG cells are able to increase their population or bIPC population by symmetric division, or differentiate into neurons via asymmetric division, while bIPCs divisions are mostly symmetric and they rarely differentiate into neurons^{83,94-98}. Roughly in the middle of the neurodevelopment process, between 17-24 post-conceptual weeks (PCW), radial glial cells undergo changes, with vRG cells losing their basal connection with the pia surface but retain their apical connection with the VZ (now referred to as truncated RG, tRG); vice versa, oRG cells maintain their connection with the pia surface, facilitating migration of newborn neurons throughout the developing cortex⁹⁹. The study of the human-specific development of the embryonic neural system has provided insight into the primate-specific complexity of the brain, characterized by a higher number of neuronal cells and a large cortical area folded in the brain. Compared to mice, human brain organogenesis is distinguished by a large SVZ and the presence of oSVZ, with the thickness of the oSVZ correlating with cortex expansion across primates⁹⁷. oRG cell proliferation seems to be the major factor responsible for the expansion of the SVZ area, indicating the increased proliferation capability of oRG cells as a human-specific key feature for the enlarged cortex. Thick oSVZ has also been observed in pallium progenitor zones, precursor regions of the telencephalon, as well as in the cerebellum, where oRG-like cells are present in humans but not in mice¹⁰⁰⁻¹⁰². The ability of oRG cells to proliferate and generate different types of neurons is the reason for the evolution of the cerebral cortex from a lissencephalic to a gyrencephalic state^{103,104}.

In the last decade, several studies on primary human embryonic tissues have applied single-cell RNA-seq (scRNA-seq) to characterized progenitor and embryonic cells at a molecular level. This has led to the discovery of new molecular markers for various cell types, including oRG cells¹⁰⁵⁻¹⁰⁷, neuroepithelial progenitors, vGR cells, IPCs, and oRG cells committed for the neurogenic or the gliogenic differentiation¹⁰⁸⁻¹¹¹, adding an accurate molecular characterization besides the already known morphological and functional characterization of progenitor cells. scRNA-seq experiments have been conducted in several regions of the human brain, including the hippocampus¹¹², striatum¹¹³, cerebellum^{114,115}, midbrain¹¹⁶, and cortex¹¹⁷⁻¹¹⁹.

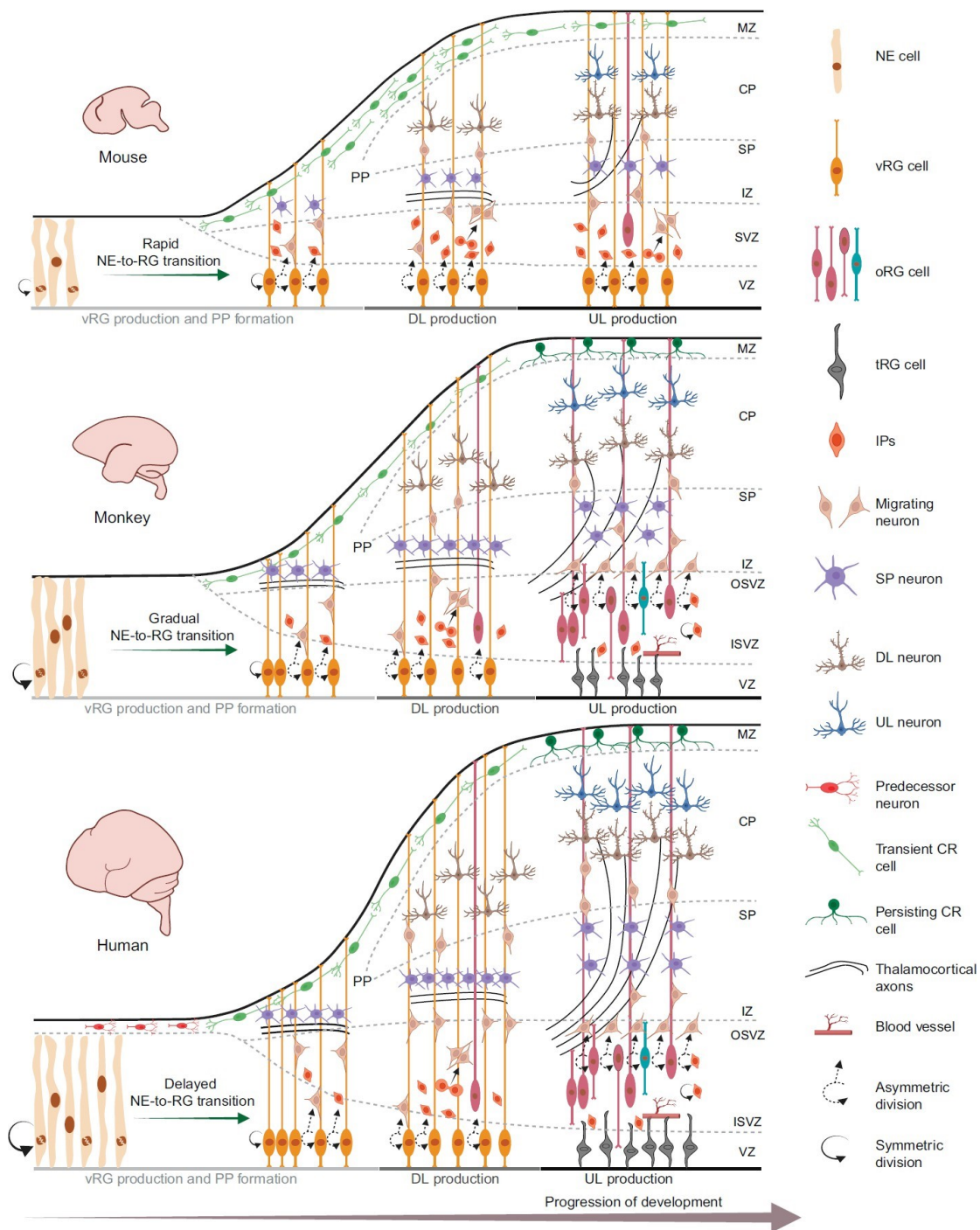


Figure 4. Cortical neurogenesis in mouse, monkey, and human. Figure adapted from Zhang et al. 2023¹²⁰. Neuroepithelial cells (NE) become radial glial (RG) cells in the ventricular zone (VZ). RG cells expand their population or contribute to neurogenesis via respectively symmetric or asymmetric division. RG cells can either differentiate directly into neurons or into intermediate progenitors, which proliferate further before differentiating into neurons. Striking differences are present in primates compared to mouse. NE cells have a longer expansion stage in primates compared to mouse¹²¹ (highlighted by the arrow thickness). Transition of NE to RG is longer in humans and monkeys compared to mice¹²². Outer subventricular zone is

increased in size in primates compared to mice^{90,93,123}, and overall, there is an increased proliferation of progenitor cells resulting in increased volume.

1.2.6 Neurogenesis and gliogenesis

1.2.6.1 Neurogenesis

Neurogenesis involves the differentiation of neural progenitor cells into neurons. The process of neuron differentiation begins early during embryonic development, following the neurulation process around 4 post-conceptual weeks (PCW), and continues until 23 PCW in most brain regions. However, in certain regions such as cerebellum or hippocampus, neurogenesis extends into the postnatal period¹²⁴. Thus, the onset and duration of embryonic neurogenesis vary depending on the specific region of the nervous system, suggesting the involvement of different regulatory mechanisms during nervous system development^{125–127}. In humans, cognitive functions such as decision-making, perception, and language, are primarily associated with and carried out by the cerebral cortex. This region exhibits significant differences between primates and rodents, which are responsible for primate-specific high-order cognitive function^{95,128,129}. Therefore, the following paragraph will focus on neurogenesis in the cortex.

The first stage of cerebral cortical development, also known as corticogenesis, involves the generation of the preplate or primordial plexiform layer, which is located between the ventricular zone and the pia membrane of the dorsal telencephalon. This region emerges around embryonic day (E) 33^{130,131} and primarily consists of early differentiated neurons known as pioneering neurons. These pioneer neurons include several populations such as Cajal-Retzius cells (CR, which are important for migration signalling), predecessor cells, and subplate cells^{83,132–134} (SP). The preplate region is highly dynamic, with post-mitotic neurons migrating inside it from the SVZ and VZ regions. Subsequently, the preplate region transforms into the cortical plate, while the intermediate zone develops above the SVZ and VZ. During this transition, CR cells migrate above the cortical plate and underneath the pia membrane, forming the marginal zone (MZ). Simultaneously, SP cells migrate beneath the preplate region and form the subplate (SP) between intermediate region and cortical plate^{83,135–137}. In the mature cortex, the MZ eventually develop into the first layer⁸³ and contains various types of cells and even dendrites from neurons located in other layers. The SP region is present only during embryonic development and plays a crucial role in neural circuit formation and axon guidance^{138–140}; a specific feature of humans and primates is the secondary expansion of the SP region¹⁴¹. Once the MZ and the SP zones are established, the upper and lower boundaries are defined, and neurons begin migrating towards the cortical plate from the VZ and SVZ, utilizing the basal projection of the oRG cells as a scaffold. Neuronal migration continues until reaching the CP region, where signalling pathways induce neurons to halt migration, differentiate, and extend dendrites and synaptic connections¹²⁹. The six cortical layers are generated by neuronal migration in an “inside-out” manner. Deep-layer (DL) neurons, located in the V and VI layers, are the first primate neurons generated¹⁴². Subsequently, the IV layer develops, with neurons receiving connections from the thalamus¹⁴³. Finally, upper-layer (UL) neurons differentiate into the layers II and III, characterized by corticocortical projections, and are primarily derived from oRG cells⁹⁰.

The capacity of precursor cells to differentiate into a wide variety of different neurons remains unknown. Several models have been proposed in an attempt to elucidate this unique ability: one model suggests an early specification of progenitor into specific neuronal subtypes during development, while another proposed that homogeneous progenitors give rise to diverse neurons during the neurogenesis stage¹⁴⁴. Recent scRNA-seq studies are beginning to provide initial evidences of the molecular mechanisms

underlying precursor differentiation. These studies suggest a process of “priming” of the RG progenitor cells for differentiation into specific neuronal subtypes prior to cell division¹¹⁹.

1.2.6.2 Gliogenesis

The process of non-neuronal cell differentiation, termed gliogenesis, begins after the neurogenesis period and continues post-natally^{129,145–147}. Non-neuronal cells account for half of the cells in the adult human brain¹⁴⁸, consisting of astrocytes, oligodendrocytes, microglia cells, and endothelial blood vessel cells. A significant portion of the human cortex is composed of astrocytes, endothelial cells, and oligodendrocytes, forming the white matter region of the cortex^{149,150}. In humans, as well as in primates, precursors of astrocytes and oligodendrocytes are generated from RG cells, particularly in the oSVZ region¹⁵¹. Single-cell studies have revealed that human oligodendrocyte progenitor cells (OPCs) differentiate from oRG cells specifically around 27 gestational weeks (GW). Before differentiating into oligodendrocytes, OPCs first enter a pre-OPC state with a molecular profile similar to oRGs. Furthermore, before differentiation, OPCs expand their population through symmetric division to compensate for the high neuron number generated during the neurogenesis¹⁵². Microglia, the most abundant immune cells resident in the brain, differentiate from primitive macrophages originating from the yolk-sac around 4 PCW. They migrate into the neural tissue and actively participate in regulating synapse and neural circuit development^{153–155}. Molecular studies at the single-cell level of microglia embryonic development have led to the discovery of a transient phagocytic cell not present in adult microglia generation. Mouse studies have identified that intermediate phagocytic microglia are associated with the myelination process and exhibit a transcription profile similar to that found in neurodegenerative diseases. This points to neurological disorders associated with disrupted microglia development during embryogenesis^{156–158}.

Once newborn neurons are generated, they must establish functional connections with their neighbouring cells. Circuit formation commences during the mid-gestation period and continues for up to a year after birth, although in some regions of the central nervous system (CNS) it persists for a longer period^{159–161}. The process entails the formation and elongation of axons and dendrites, followed by synapse formation and pruning (a mechanism that removes the unnecessary synaptic connections), and finally, axon myelination. Studies conducted at single-cell resolution have demonstrated that this process is conserved between primates and humans, although it takes longer in humans¹⁶². During circuit development in the cortex, immature circuits are generated by spontaneous electrical activity of neurons in the subplate region, as evidenced by groundbreaking work in cats¹⁶³. Subplate neurons play a crucial role in early embryonic synaptogenesis and spontaneous circuit formation but are replaced during development and are not present in the adult cortex¹⁶⁴. These neurons are more mature compared to those in the cortical plate¹¹⁸, and can induce the migration of cortical neurons by stimulating them with synapses¹⁶⁵; the significant role of subplate neurons during circuit formation is underscored by the enlarged subplate region observed in the primate cortex, suggesting enhanced formation of functional connections in primates¹⁶⁴. In humans, dendritic spines begin to develop during the second trimester, experiencing a burst of growth between 3 to 24 months after birth, depending on the specific cortical region¹⁶⁶. By one year after birth, circuits enter the maturation stage, characterized by proper synaptic pruning and morphological changes¹⁶⁷.

1.2.7 *In vitro* modeling of human neuronal differentiation and brain organogenesis

Neuroscientific discoveries have relied on direct observation of phenomena in alternative models to shed light on the human-specific neural embryonic developmental process. Until recent years, the study of human brain organogenesis was severely limited by the absence of primary tissue. However, the

development of protocols and methods for the generation of neuronal cells from human-derived PSCs has led to the discovery of key aspects of human CNS organogenesis^{168–171}. Since the early 1970s, scientists have been able to isolate and cultivate mammalian embryonic brain tissues *in vitro*, forming structures resembling *in vivo* areas of the brain such as the cortex or hippocampus, as well as neuromuscular junctions^{172–174}. With the advent of human pluripotent stem cells, both isolated from embryo^{47,56} (hESCs) or reprogrammed from somatic cells¹⁷⁵ (hiPSCs), protocols for the generation of neurons *in vitro* were developed, leveraging small molecules and growth factors in the media¹⁷⁶. Neuron differentiation of hPSCs was achieved in the dish, creating two-dimensional (2D) cultures of different types of neuronal cells¹⁷⁷. However, despite being a homogeneous model suitable for large-scale studies, 2D cultures failed to recapitulate the subtle cellular interactions with other cells or the cellular matrix important during embryonic development, thus generating *in vitro* artifacts^{178,179}. Furthermore, 2D cultures cannot be maintained for the amount of time necessary for processes such as circuit formation and gliogenesis, which occur later during human embryonic and neonatal development. In the early 2010s, protocols were developed for generating three-dimensional (3D) hPS cell agglomerates that self-organize into neural structures and tissues, mimicking interactions and tissue architecture of embryonic structures; these cell cultures were called organoids¹⁸⁰. The first neural organoids generated were optic cup organoids¹⁸¹. Protocols for generating brain organoids are generally divided into unguided and guided differentiation protocols. Unguided protocols rely on the self-patterning and self-organization characteristics of hPSCs, with cells embedded into Matrigel or other extracellular matrices (ECMs) able to differentiate into various cells and neuronal structures like forebrain, hindbrain, retina, and non-ectodermal cells: due to the unguided nature of the differentiation, the variability of cell composition between different experiments is elevated, making it difficult to mimic subtle disease phenotypes. Furthermore, spontaneous cell differentiation can lead to false-differentiation into non-neuronal cells¹⁸². On the other hand, guided protocols leverage the discovery of pathways and signalling molecules driving neural differentiation, for example inducing the inhibition of SMAD for neuroectodermal induction, and subsequently administering specific morphogens and molecules for region-specific organoids, achieving more reproducible results within experiments^{183–186}. Over the last decade, protocols for ECM-free organoid development, as well as for post-natal developmental stages, have been developed, expanding the possibilities to use organoid technology to study physiological and pathological aspects of neural development^{187–190}.

1.3 X-chromosome Inactivation

In eutherian mammals, males and females are genetically distinct due to the presence of sex chromosomes: XX in female and XY in male. Genes transcribed by the sex chromosomes initiate sex-specific development programs that constitute the basis of the sexual dimorphism between males and female¹⁹¹. The X and Y chromosomes have evolved more than 160 million years from the same pair of autosomes. A progressive degeneration of the Y chromosome has led to a loss of most of its genetic content. The X chromosome is 160 Mb long and encodes more than 800 protein-coding genes, while the Y chromosome is 60 Mb long and encodes less than 100 genes^{192–196}. The discovery of sex as a genetic hereditary factor began in 1905 when Dr. Nettie M. Stevens described the X and Y chromosomes responsible for the sex determination in mealworms¹⁹⁷. Based on studies on *Drosophila*, in 1947 Muller hypothesized that the expression of genes from the common sex chromosome should be balanced between males and females, introducing the dosage compensation theory¹⁹⁸. Soon after, Barr and Bertram first described a nuclear structure identified in neurons of a female cat, not present in male neurons, termed “nucleolar satellite”. They provided the first empirical evidence of the (inactivated) additional X chromosome in female mammals, with its peculiar morphological characteristics resulting from dosage compensation¹⁹⁹. In 1960, Ohno re-proposed that the visible nucleolar satellite was the

additional X-chromosome present in female somatic cells, identifying the same nuclear structure in mice female cells²⁰⁰. The dosage compensation theory was embraced by Mary Lyon to explain the observed nuclear structure of the additional X-chromosome. She postulated that in females, one of the X-chromosomes should be inactivated to balance the expression of X-linked genes between males and females. This inactivation should occur early during embryonic development, and once inactivated, the same chromosome remains inactive in the daughter cells. This marked the first enunciation of the X-Chromosome Inactivation process²⁰¹. The pioneering work of Mary Lyon was inspired by calico mice and cats, and the theory was later developed by Ohno, suggesting that the process of dosage compensation between males and females could occur via two mechanisms: the upregulation of X-linked genes in males, so that the X-linked gene expression would be double the normal expression of the X-chromosome in females or the inactivation of one of the two X-chromosomes in females, so that the X-linked gene dosage would be the same between males and females²⁰². After more than 60 years of empirical evidence, the mechanism of X-Chromosome Inactivation (XCI) is widely accepted as a general mechanism for the random inactivation of one of the two X-chromosomes in female somatic cells, leading to dosage compensation of X-linked genes between males and females in mammals.

1.3.1 Sexual dimorphism in mammals

Sexual dimorphism is the phenomenon that allows distinguishing the sex of a species by their phenotype, known as mammalian sexual dimorphism. Phenotypic differences between males and females result from genotypic differences in the sex chromosomes, which determine the phenotype of the gonads. These gonadal sexual differences culminate in the release of sex-specific hormones, which reach virtually any cell in the body, activating a sex-specific gene expression cascade that results in the sexually dimorphic phenotypes^{203,204}. In mammals, sexual dimorphisms are associated with visible phenotypic traits of the species, including body size, with adult males generally being larger than females²⁰⁵, while for some species females are larger than males²⁰⁶. Recent advances suggest that differences in body size are not as pronounced as previously claimed, with new estimates indicating that males are not larger than females in most mammals²⁰⁷. Other traits such as body and skeletal composition show differences in muscle distribution and bone growth^{208,209}, as well as variation in vocalization, body marking, and pelage²¹⁰.

Sexual dimorphisms are the phenotypical results of sex-biased gene expression differences across the body²¹¹. Several studies have investigated sex-specific gene expression programs in humans and different mammals, in adult individuals and during development, at organ-specific resolution^{212,213}. The number of sex-biased genes varies depending on the mammalian species, with humans having 37.5% of genes being sexually biased in at least one tissue. However, this percentage drops for mice (5.9%) and other mammalian species tested, indicating that the effect of sex on the transcriptome is widespread but not strong in all species. Most of the sex-biased genes identified are species- and tissue-specific^{212,214}. Few genes were identified as having sex-biased expression across all tissues, and a small set of genes showed a conservation of sex-bias expression in more than two species. These genes are mostly associated with sex chromosomes, such as X gametologs with a counterpart on the Y chromosome, and long non-coding RNAs (lncRNAs) involved in the X chromosome inactivation (XCI) process^{212,214}. Tissue-specific sex-biased genes evolve with distinct mechanisms, leading to organ-specific sexual dimorphisms in organ function as well as in diseases associated with the organ²¹⁵⁻²¹⁸.

1.3.2 X-chromosome Inactivation as a mechanism of dosage compensation in mammals

In eutherian mammals, the mechanism to achieve dosage-compensation of X-linked genes between sexes is called the X-Chromosome Inactivation process (XCI), which leads to the random inactivation of one of the two X-chromosomes present in female cells²⁰¹. The inactivation process occurs early during

embryonic development and induces several changes in the structure and nuclear localization of the inactive X-chromosome, achieving transcriptional silencing²¹⁹. The mechanisms and modalities by which the additional X-chromosome is silenced are specific and different for each species²²⁰. Differences can be found in the early stages of XCI: for example, marsupials are characterized by rapid inactivation of the paternal imprinted X chromosome in the early stages of development, with a conserved process leveraging the lncRNA *Rsx*^{221,222}. In eutherian mammals, lncRNAs play a crucial role in the silencing process, but the initiation of the XCI differs depending on the species, as studies have shown in mouse, rabbit, cynomolgus monkey, and human, indicating an independent adaptation of the XCI process and differing lncRNAs expression patterns and actions^{223–225}. One example is the presence of imprinted XCI in both marsupials and mouse: while in marsupials, the paternal X-chromosome is always imprinted and inactivated, in mice, XCI of the imprinted paternal X-chromosome is established early after fertilization (four-cell stage), followed by X-Chromosome Reactivation (XCR) in the blastocyst stage in the Inner Cell Mass, with subsequent establishment of dosage compensation via random XCI after implantation²²⁶. In rabbits, cynomolgus monkeys, and humans, XCI initiates at later developmental stages and does not include any imprinting²²⁷. In humans and cynomolgus monkeys, the pre-implantation embryo possesses two actively transcribed X chromosomes, and random XCI occurs after implantation^{224,228,229}. Dosage compensation of X-linked genes via XCI is crucial for proper embryonic development: the distorted sex ratio towards male newborns after *in vitro* fertilization (IVF) procedures in humans and mice has been linked to improper XCI in female embryos²³⁰, while absence of dosage compensation and biallelic expression of X-linked genes, caused by deletion of crucial lncRNA involved in the XCI process specifically in the paternal X-chromosome, leads to female-specific embryonic lethality due to extraembryonic tissue degeneration^{231,232}.

1.3.3 lncRNA *XIST* expression during early development (initiation of XCI)

In eutherian mammals, the primary orchestrator of the XCI process is a lncRNA called X-inactive specific transcript (*XIST*)^{233–235}. First identified in mouse and human, this gene is located in the X-Inactivation Centre (XIC) on the X-chromosome, and its expression is sufficient to induce the inactivation of the chromosome from which it is transcribed in cis²³⁶. The *XIST* gene was discovered in the early 1990s, marking one of the first lncRNAs to be identified. It is transcribed into RNA, capped, and polyadenylated, but it remains in the nucleus where it accumulates on the X-chromosome from which it is transcribed, effectively “coating” the X-chromosome and leading to its transcriptional inactivation^{237–239}. The mechanism of this coating is still not elucidated, but it involves *XIST* first accumulating in X-chromosome regions known as “entry sites” closer to the *XIST* transcriptional site, then spreading to the entire X-chromosome²⁴⁰. Studies utilizing transgenic approaches to delete the *Xist* gene in Embryonic Stem Cells and embryos have demonstrated that in the absence of *XIST*, cells fail to induce XCI. Conversely, ectopic expression of *XIST* from autosomes induces cis-transcriptional silencing of the expressing autosome, highlighting the essential role of *XIST* expression as an early event for proper XCI^{241–245}.

The processed *XIST* transcript is a large RNA molecule, 17 kb in humans and 15 kb in mice, composed of several repetitive sequences named from A to F repeats. These repeats are highly conserved in mammals, although there are differences in the number of repetitive motifs among different species^{239,246–248}. *XIST* serves two important functions during XCI: (i) it acts as a scaffold for several proteins and complexes involved in the XCI process, and (ii) it guides them to the X-chromosome that needs to be inactivated. The diverse functions of *XIST* during XCI, including binding and coating the X-chromosome and recruiting repressive protein complexes, are mediated by distinct transcript domains

identified through genetic and biochemical approaches, partially overlapping with the repeat domains²⁴⁹. The A repeat of the *XIST* transcript is the most conserved among species and has been identified as the key element triggering transcriptional silencing. It is located at the 5' end of the transcript and consists of 7.5 copies of a 26 nucleotides sequence. Deletion of this domain leads to improper X-chromosome silencing and impaired fetal development, as well as defective transcriptional regulation of *XIST*^{249–251}. One of the key proteins binding to the A repeat domain is SPEN (also called SHARP, SMRT/HDAC1-associated repressor protein), binding the *XIST* RNA through its N-terminal RNA recognition motifs (RRMs). SPEN is also characterized by a C-terminal Spen paralog and ortholog domain (SPOC). Thanks to the SPOC domain, SPEN is able to recruit SMRT, which in turn initiates transcriptional silencing by recruiting HDAC3 and inducing deacetylation of most of the X-linked genes^{252–257}. The complexity of *XIST*'s central role in XCI is also reflected by the multi-functional role of its binding proteins, such as SPEN. Although necessary for HDAC3 recruitment, it has been demonstrated that SPEN can bind to several other proteins involved in chromatin remodelling and epigenetic silencing, acting as a scaffold for assembling multi-protein repressive complexes^{258,259}. Repeats B and C of the *XIST* transcript function to recruit the Polycomb repressive complex 1 (PRC1), which subsequently recruits PRC2, allowing for the deposition or repressive chromatin marks on the genes located on the inactive X-chromosome^{257,260,261}. The *XIST* RNA Polycomb Interaction Domain consisting of the B repeats and parts of the C repeats (XR-PID), mediates the binding of heterogeneous nuclear ribonucleoprotein K (hnRNPK), which in turn recruits PRC1^{260,262}. PRC1 is then responsible for recruiting SmcHD1 and PRC2, as confirmed by proteomic studies that did not find any direct interaction between *XIST* and PRC2^{263–265}. Repressive histone marks H3K27me3 and H2AK119ub, as well as rearrangement of chromatin structure, are obtained through these protein complexes, indicating a role in maintaining the silencing of the X-chromosome rather than initiating it²⁶⁶. Genetic deletion studies have demonstrated that many of *XIST*'s functions are redundant, often involving different domains in similar functions. The localization of *XIST* RNA on the X-chromosome is partially mediated by the C, D, and F repeats, which bind to several proteins and anchor *XIST* RNA to the DNA through redundant processes^{249,267–274}. One example is the Cip-1 interacting zinc finger protein 1 (CIZ1), a protein associated with the nuclear matrix and binding the E repeat region of *XIST*. CIZ1 binds to X-chromosome DNA through its three DNA-binding zinc finger motifs and to *XIST* RNA through the C-terminal nuclear matrix-anchoring domain MH3 domain (matrin 3 homologous domain)^{264,270,271,275,276}. In addition to this, several other proteins have been shown to bind different regions of *XIST* to mediate their function. hnRNP U (also known as SAF-A) plays a key role in anchoring *XIST* RNA to the nuclear matrix by binding to the lncRNA transcript at exon 1 and 7^{269,277}. The *XIST* A repeat interacts with Wilms tumor-1 associated protein (WTAP), which in turn binds to RNA-binding motif protein 15 (RBM15), a part of the protein complex responsible for the deposition of N⁶-adenosine (m⁶A) RNA methylation^{264,278}. m⁶A is the most abundant RNA modification and it is involved in several post-transcriptional pathways and functions through the recognition and binding of m⁶A reader proteins, which trigger downstream effects on the RNA function and metabolism²⁷⁹. Upon recruitment of the m⁶A RNA methylation machinery METTL3 (abbreviation for Methyltransferase 3, N6-Adenosine-Methyltransferase Complex Catalytic Subunit), 78 adenosine residues are methylated in the *XIST* transcript, suggesting a key role of the m⁶A residuals in modulating *XIST* functions during XCI. In line with that, it has been shown that deletion of the YTH domain-containing protein 1 (YTHDC1) reader protein results in impaired XCI, while inducing YTHDC1 binding to *XIST* in the absence of functional m⁶A methylation machinery leads to a proper XCI process²⁸⁰.

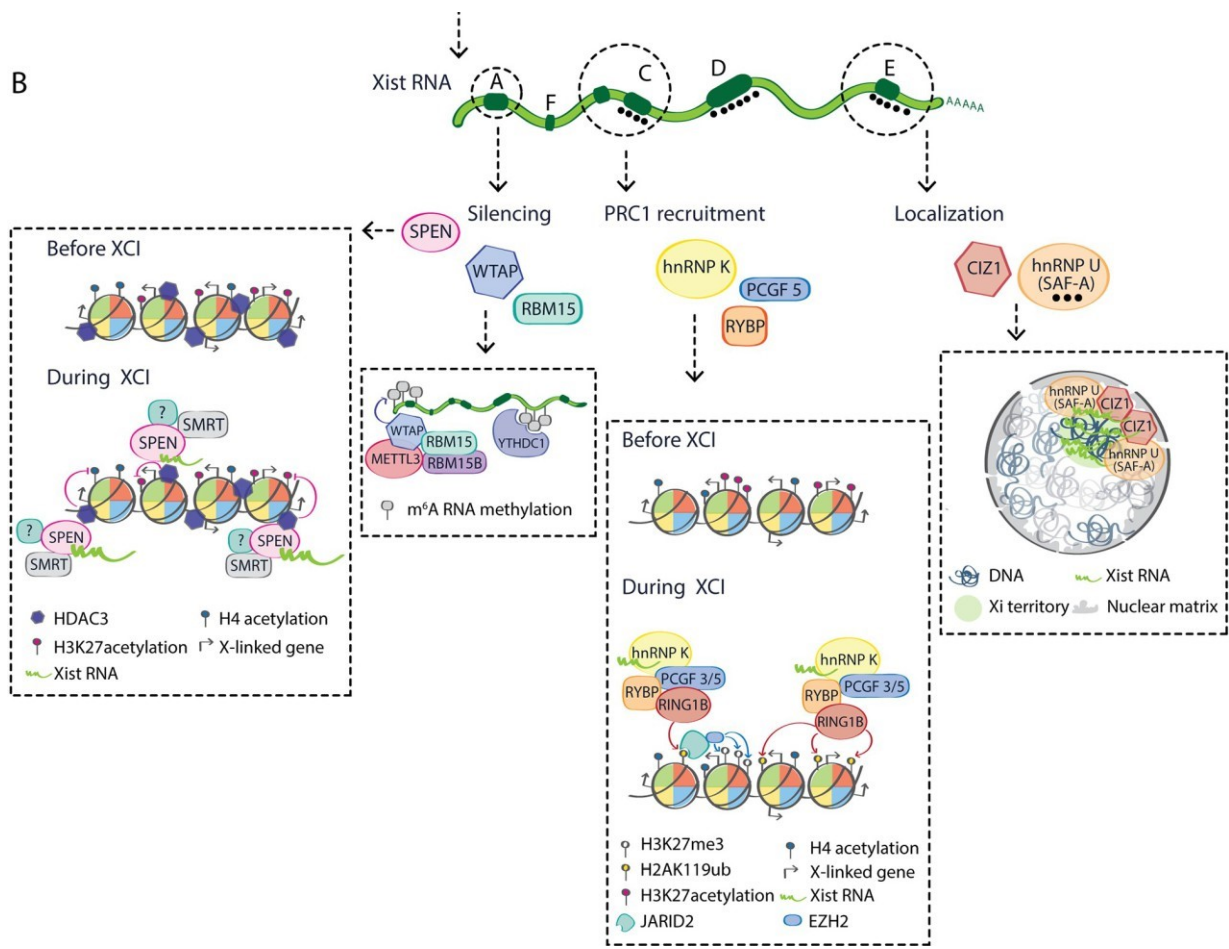


Figure 5. *XIST* RNA and its interactors mediate different functions during XCI. Figure adapted from Loda and Heard 2019²⁸¹. Schematic representation of *XIST* RNA domains and interactor proteins that mediate different roles during XCI.

1.3.4 Initiation of XCI in humans: the dual role of *XIST* and *XACT*

The lncRNA *XIST* is a complex and multi-functional molecule capable of initiating and maintaining the transcriptional inactivation of one X-chromosome. Studies on the X-Chromosome Inactivation (XCI) process in humans and mice have revealed key differences between the two species, particularly in the initiation of the XCI, emphasizing the importance of using human-specific models for the analysis of XCI. In mice, both X-chromosomes are active after fertilization due to gene reactivation induced during the one to two-cell stage of the zygote. Imprinting of the paternal X-chromosome occurs at the four to eight-cell stage, leading to its inactivation until the late blastocyst stage, where gene expression from both parental X-chromosome is reactivated, and random X-chromosome inactivation is established in epiblast cells^{282–286}. In human pre-implantation embryos, imprinted XCI does not occur. Instead, studies utilizing RNA-FISH and scRNA-Seq in human embryo have demonstrated active *XIST* expression from both X-chromosomes in female cells and from the single X-chromosome in male embryos. Both coated X-chromosomes in human pre-implantation embryos are transcriptionally active, revealing significant transcriptional difference of the *XIST*-coated X-chromosomes between mouse and human embryo^{224,227,287}. Based on scRNA-Seq studies on embryo, Petropoulos and colleagues²²⁴ identified a

gradual decrease in expression from X-chromosomes expressing *XIST* starting from the morula stage. This suggests an initial dosage compensation process occurring in human pre-implantation embryos, referred to as the dampening of X-chromosome. However, this theory is still debated due to contradictory results obtained from the reanalysis of the same dataset. But confirmation of the dampening process of *XIST* coated X-chromosomes came from work in naïve stem cells^{288–290}. Nevertheless, the dual behaviour of the coated X-chromosome, with active transcription before implantation and repression afterwards, remains elusive to explanation. Analysis of human embryos and naïve cells has shown a more dispersed cloud of *XIST* in cells expressing alleles from both X-chromosomes compared to those with random XCI, suggesting a more relaxed coating of the X-chromosomes in the pre-implantation stage, and providing a potential explanation for the absence of the complete gene repression^{287,289}. Moreover, in naïve pluripotent cells and early embryonic stages, biallelic *XIST* expression is associated with the expression of a human-specific lncRNA, X-active specific transcript (*XACT*). Based on the co-expression analysis of these two lncRNAs and their mutual location within the X-chromosome, it has been postulated that *XACT* interferes with the *XIST*'s repressive action during pre-implantation stages by maintaining a dispersed *XIST* coating²⁸⁷. However, studies on hPSCs *in vitro* models have described that the role of *XACT* is dispensable in cells where the inactive X-chromosome is artificially re-activated due to loss of *XIST* coating and repressive histone marks, a process called erosion of the X-chromosome²⁹¹. This raises questions about its antagonistic role with *XIST*²⁸⁷.

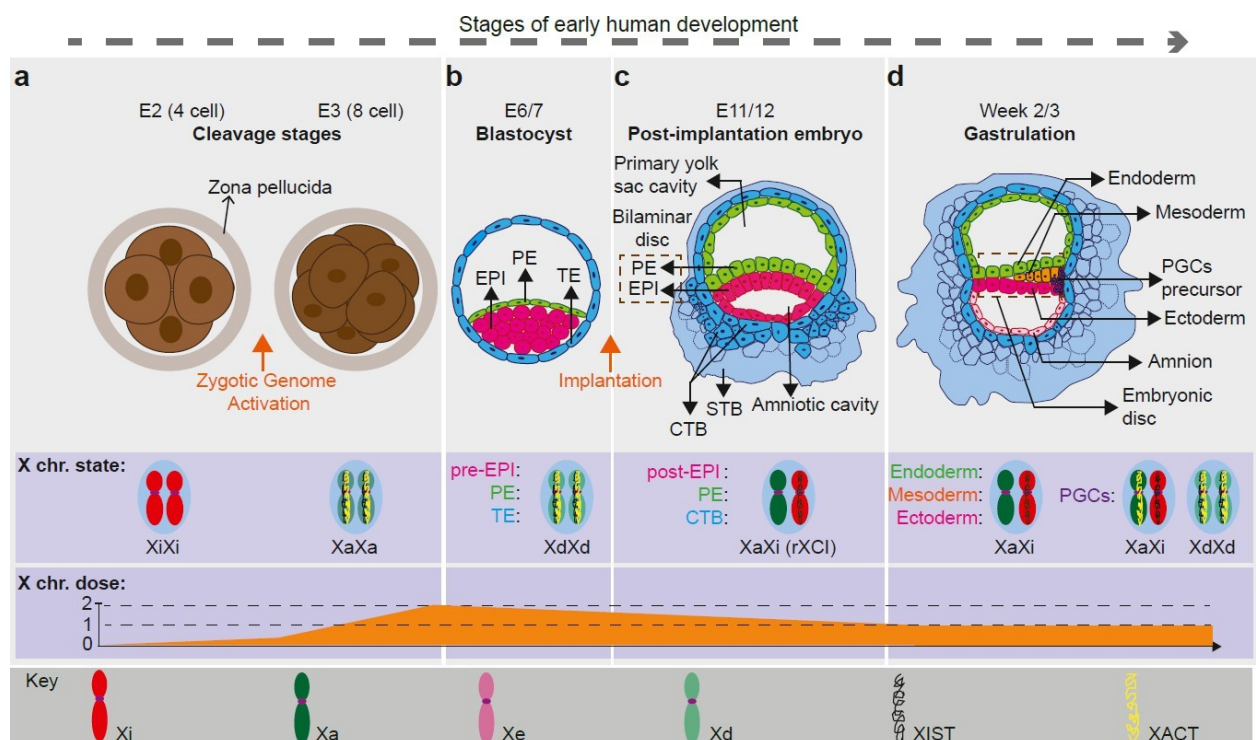


Figure 7. X-chromosome states during early human development. Figure adapted from Khan and Theunissen, 2023⁷⁶. Schematic cartoon representing the early human embryonic development and related *XIST* and *XACT* expression, X-chromosome state, and dosage compensation status of X-linked gene expression.

1.3.5 Role of inactive X-chromosome in sexual dimorphism: escapee genes

Despite the finely tuned process of X-chromosome transcriptional inactivation, the inactive X-chromosome actively contributes to sex differences between males and females, thanks to a subset of genes that are expressed despite the global epigenetic silencing. These genes are referred to as escapee genes or escapees²⁹². In female humans, genes identified as biallelically expressed from both X-chromosomes comprise more than 20%^{293,294}. Escape genes have also been identified in mice, although at much lower numbers, making the mouse model of questionable value for the prediction of escape genes in humans²⁹⁵. Escapees are capable of bypassing the tight epigenetic repression established on the inactive X-chromosome, acquiring an epigenetic signature similar to genes located on the active X-chromosome. The expression levels from the inactive X-chromosome (Xi) relative to the active X-chromosome (Xa) differ depending on the genes. Some genes are expressed at comparable levels between Xi and Xa, while others have a lower expression from Xi compared to Xa. Furthermore, escapees are classified based on their ability to “escape” XCI consistently in all cell types, different individuals, and during different developmental stages and in different tissues. These are referred to as “constitutive escapees”, suggesting that they escape inactivation since its establishment and are not silenced²⁸⁴. Conversely, a subset of escape genes has been identified as variably escaping from the inactive X-chromosome. These genes are biallelically expressed specifically in certain tissues or cells, individuals, or during specific differentiation lineages or times, indicating the possibility of a dynamic process of gene expression reactivation upon transcriptional silencing, or late silencing during development²⁹⁶.

1.3.6 Methods for investigating escape genes

Due to the challenges in obtaining reliable information on human escapees using the mouse model, several studies have attempted to establish a consensus for human escape genes based on various technical and methodological approaches. A common definition of an escape gene is one whose expression from the Xi is greater than 10% of its expression from the Xa²⁹². Early studies on the expression of genes from the inactive X-chromosome were conducted in mouse/human hybrid cells, where one X-chromosome is of human and the other is of murine origin. By isolating hybrids containing the Xi derived from humans and comparing their genome-wide expression of X-linked genes to hybrids containing the human Xa, researchers found that approximately 10-15% of X-linked genes were expressed from the Xi in hybrid cells. Although the altered epigenetic landscape due to the failed *XIST* coating of the Xi in hybrid cell models may influence these findings, they were partially confirmed using complementary approaches^{292,297,298}. Another valuable approach for identifying escapee genes is to compare male-female X-linked gene expression to detect female-biases genes^{212,294}. Generally, escape genes can have higher expression in female cells due to biallelic expression compared to male²⁹⁹. However, this approach is susceptible to false negative results for genes with low expression from the Xi and false positives due to female-biased expression resulting from sex hormones activity. To address these issues, transcriptomic data can be analysed based on the monoallelic or biallelic expression of polymorphism found in the coding regions of the X-linked genome, allowing for the simultaneous tracking of the allelic expression of several candidate genes. However, this approach is heavily limited by the presence of expressed polymorphisms and the mosaic landscape of the random XCI, where, on average, both X-chromosomes are active in a population of cells. Transgenic approaches have been used to achieve controlled XCI of specifically one X-chromosome, improving the quality of transcriptomic data^{257,300}. In 2017, Tukianer and colleagues identified a female donor in the GTEx dataset with a completely skewed X-chromosome³⁰¹. This rare discovery provided an opportunity to investigate tissue-specific mono- and biallelic expression of X-linked genes with expressed polymorphism. Out of over

1000 X-linked genes analyzed, only 186 met the expression threshold criteria for the analysis. Among the analysed genes, 23% exhibited tissue-specific biallelic expression²⁹⁴. The limited amount of analyzed genes can be overcome by increasing the number of transcriptomes derived from individuals. scRNA-seq experiments on human tissues can be employed to study the XCI status. Despite limitation due to the technology, such as cell doublets³⁰², and missing consensus for gene expression threshold, scRNA-seq data from the GTEx dataset have been successfully analyzed and the consensus of almost 100 X-linked genes were defined. X-Chromosome Inactivation heavily influences the epigenetic landscape of the transcriptionally inactive X-chromosome, leading to the deposition of repressive histone marks and DNA modifications. However, escape genes have distinct epigenetic marks compared to genes subject to XCI, providing the opportunity to detect escape genes in epigenetic analysis on the X-chromosome. A well-established analysis based on changes in repressive epigenetic marks is the DNA methylation analysis of CpG islands in the promoter regions of X-linked genes: by comparing models lacking Xi and containing both Xa and Xi (for example male vs female), similar methylation levels can be associated with active transcription from both Xi and Xa³⁰³⁻³⁰⁵. Other approaches involve the use of additional chromatin marks³⁰³, as well as RNA Pol II accessibility and binding³⁰⁶. In addition to histone marks analysis, accessibility analysis of the DNA using ATAC-seq (Assay for Transposase- Accessible Chromatin) technology is often employed to define actively transcribed genes as those with an open promoter compared to genes subject to XCI, which typically have a more inaccessible and closed promoter region³⁰⁷. Finally, fluorescent in situ hybridization technologies allow for the direct visualization of gene expression from Xa and Xi. By using probes recognising intronic sequences of genes, unspliced pre-mRNA can be detected retained at the loci of RNA synthesis³⁰⁸. However, high throughput approaches are laborious by single gene visualization with RNA-FISH, and expression level of genes must be high enough to be visualized.

1.3.7 Role of escape genes in disease phenotype

The presence of escape genes on the inactive X-chromosome highlights the importance of the inactive X-chromosome in the dosage compensation and gene expression portfolio in female cells. This is also reflected in diseases where the number of inactive X-chromosomes changes. For example, both clinical phenotypes of Turner syndrome (45X) and Klinefelter syndrome (47XXY) are linked with alteration in dosage compensation of escape genes^{299,309}. For instance, the *SHOX* escape gene has recently been associated with body height in these syndromes: short stature in Turner syndrome and tall stature in Klinefelter syndrome^{310,311}. Out of the 199 reported X-linked intellectual disability syndromes (ID), 162 X-linked genes have been associated with the ID³¹², including several escape genes such as *DDX3X*³¹³, *USP9X*³¹³, *KDM5C*, and *KDM6A*^{314,315}. Due to hemizyosity of X-linked genes in males, loss-of-function mutations lead to a more severe phenotype compared to female carriers, often resulting in early lethality. The same escape genes mentioned above play an important role in cancer biology, with mutations associated with higher tumour progression, especially in males³¹⁶⁻³²⁰. The higher incidence of cancers in males can be partially explained by the EXITS theory, which postulates that females have a lower cancer incidence due to the biallelic expression of tumour suppressor escape genes³¹⁸. Sex-based differences are also frequently observed in immune responses, where males are overall more susceptible to infections, while females have an increased incidence of autoimmune diseases^{321,322}. One of the genes involved in infectious response, *TLR7*, has been identified as a variable escape gene and can differentially be expressed from the inactive X-chromosome in different individuals^{323,324}. This is a possible explanation for the sex bias of infectious disease, but also explain individually variable immune response.

1.4 Neurodevelopmental disorders

During the development of the nervous system, functional and morphological aberrations can result in neurodevelopmental disorders (NDDs). These conditions typically manifest either congenitally or during early postnatal life, although there are exceptions with a later-onset of disease^{325–328}. NDDs are characterized by a diverse range of phenotypes³²⁹. Patients can present with changes in motor ability, speech and language development, learning abilities, impairment of attention and activity, and communication. These clinical phenotypes often co-occur and are not mutually exclusive, contributing to impairment in social, educational life, and overall wellbeing of the patients^{329–333}. Impairments in brain function associated with NDD primarily affect general cognitive abilities³³⁴. The classification of NDDs encompasses a broad spectrum of conditions, including autism spectrum disorder (ASD or autism), attention/deficit hyperactivity disorder (ADHD), learning and communication disorders, intellectual disabilities, syndromes resulting from genetic aberrations (such as Rett syndrome), schizophrenia, obsessive/compulsive disorders, fetal alcohol spectrum disorders, and congenital brain abnormalities^{335,336}. The etiology of NDD is complex, involving genetic inheritance and multifactorial causes related to interactions between genetic and environmental factors (nature-nurture interactions)^{337–341}.

1.4.1 Sex differences during neurodevelopmental process

The term “sex” refers to distinct yet interconnected biological attributes primarily associated with reproduction, encompassing genetics, hormones, gonadal and genital features, as well as anatomical and physiological components^{342–345}. In humans and mammals, sex is determined by the presence of the sex chromosomes X and Y, with females having XX and male having XY karyotypes. Less common variations in sex chromosome status include sex chromosome aneuploidies. The presence of the Y chromosome is pivotal in determining male sex, as it carries the *SRY* gene responsible for testis development. In the absence of the SRY protein, individuals develop ovaries as part of the default developmental program. Testes produce testosterone, which influences various aspects of development, including the formation of gonadal and reproductive organs, brain development, and overall physiological processes. Hormonal fluctuations during puberty further contribute to physiological differences that persist throughout life. Genetic and chromosomal disparities between the sexes involve mechanisms such as genomic imprinting, mosaicism, and X-inactivation, as well as the presence of escape genes. These complexities underscore the limitation of binary sex labelling at birth, highlighting the intricate nature of sex-related biological variables present in newborn.

Brain development is profoundly influenced by gonadal sex hormones, which are sex-specific features resulting from genetic sex differences^{346–348}. The male-specific *SRY* gene triggers the release of fetal testosterone during the second trimester of pregnancy, leading to the masculinization of the brain³⁴⁷. Additionally, gene expression exhibits strong sex bias in nearly all tissues, influenced by the combined effects of sex chromosomes (including genes escaping X-chromosome inactivation), different deposition of epigenetic marks on autosomes, and the action of transcription factors regulated by sex hormones²¹². While the overall structure and organization of the human brain do not show clear differences between males and females, there is considerable inter-sex variability in brain features³⁴⁹. The literature presents conflicting findings regarding male/female brain differences, with some studies reporting larger volumes in certain brain regions in males, such as the limbic and temporal regions, while others indicate larger volumes in females, particularly in the cingulate and prefrontal regions^{350,351}. Functional disparities between the sexes include females exhibiting better communication abilities and analytical-intuitive thinking, whereas males tend to excel in motor tasks and perception abilities³⁵¹.

1.4.2 Sex bias in neurodevelopmental disorders

Males exhibit a higher prevalence of all neurodevelopmental conditions compared to females, with a sex ratio ranging between 1.2:1 and 4:1. Large studies conducted in children have reported an overall rate of neurodevelopmental disorders (including ADHD, autism, speech and learning disorders, and intellectual disabilities) of 12.6% in females and 22.7% in males³⁵². Sex differences are more prominent in children than in adult, particularly in autism, ADHD, and tic disorders^{353–357}, and sex ratios vary depending on the phenotype of the specific disorder. For instance, severe forms of autism, referred to as profound autism, do not exhibit any sex bias³⁵⁸.

Behavioural differences between male and female have been observed in various NDD, such as autism, ADHD, communication and learning disorders, motor disorders, and intellectual disabilities³⁵⁹. Additionally, sexual differences can impact the response to therapies³⁵⁹.

Various theories have attempted to elucidate the sex bias observed between males and females in neurodevelopmental disorders. One theory proposed a female protective effect against autism, suggesting that females inherit a greater number of genetic variants predisposing to autism than males, yet they are generally protected from developing autism, possibly due to factors such as having an additional X chromosome^{360,361}. Conversely, sex chromosome aneuploidies (SCAs) have been associated with a higher incidence of several NDD, implying that disruption in gene dosage compensation of sexual genes may contribute to the development of these NDD³⁶². Genes that escape XCI play a particular role in the pathology of gonosomal aneuploidies³⁶³. Moreover, skewed X-chromosome inactivation has been implicated in differences in the phenotypic expression of certain disorders in females, such as Fragile X syndrome (FXS)³⁶⁴, underscoring the significance of X-chromosomal gene expression in phenotype manifestation. Additionally, autosomal genes have been found to exhibit sex-biased expression both *in vivo* in disorders like ADHD and autism^{365,366}, as well as in animal models³⁶⁷.

1.4.3 Opitz/BBB G Syndrome

Opitz/BBB G syndrome is a congenital neurodevelopmental disorder with heterogeneous genetic causes. The incidence of Opitz disease ranges from 1 in 50000 to 1 in 100000, although epidemiological data are limited. Initially, Opitz BBB and G syndromes were reported as two distinct diseases, distinguished by the presence of laryngeal malformation in the G syndrome but not in the BBB syndrome. However, segregation analysis of affected families revealed the co-segregation of these diseases, indicating a unique disorder associated with the same gene but with a wide range of different phenotypes. Opitz syndrome has been linked to an X-linked recessive form, characterised by mutations in the *MIDI* gene (X-linked Opitz syndrome, XLOS)^{368,369}. In 1995, Robin and colleagues³⁶⁹ identified an additional GBBB2 disease associated with an autosomal dominant variant, specifically located in the 22q11.2 chromosomal locus. The autosomal form of Opitz syndrome shares several phenotypic features with the X-linked disease, but distinct phenotypic differences and the association with the 22q11.2 locus led to the establishment of distinct diagnoses of BBBG- syndrome and BBBG-like disease. Subsequently, autosomal GBBB2 was found to be associated with mutations in the Sperm Antigen with Calponin Homology and Coiled-Coil Domains (*SPECCIL*) gene, exhibiting characteristics more similar to the TBHS-1 syndrome (Teebi hypertelorism syndrome-1, MIM145420), and they are now considered the same condition³⁶⁹. In the early 2000 the X-chromosomal gene *MIDI* was discovered to be associated with the X-linked Opitz form³⁷⁰.

Clinical manifestations of XLOS include characteristic facial features, with hypertelorism being the most common (reported in more than 95% of cases), followed by cleft lip/palate (reported in 50% of population), and other common attributes like a broad nasal bridge and tip, and a prominent forehead.

Body malformations are also common, with more than 80% of patients affected by hypospadias, along with other common malformations including laryngeal cleft (in 50% of patients), esophageal reflux, and associated problems (in 30% of patients), heart malformations (in 20%), as well as ectopic or imperforated anus (20%). One-third of affected males report Intellectual Disabilities (ID) and/or delays in cognitive development, with brain abnormalities often reported, such as reduction in volume and dimension (hypoplasia) of the corpus callosum and the vermis. Males are generally more affected than females, with females only exhibiting hypertelorism in more than 90% of cases, and rarely other malformations or features^{369,371–376}.

XLOS is generally characterized by alterations in the midline of the body during embryonic development, attributed to mutations in the midline 1 (*MID1*) gene, a master regulator of midline structure development.

1.5 MID1

The *Midline1* gene (*MID1*, or *TRIM18*) is an X-linked gene that encodes an E3 ubiquitin ligase protein, belonging to the tripartite motif (TRIM) family. The *MID1* gene's location is similar in mice and humans, but not precisely conserved: the *Midl* orthologues gene in the mouse is situated in the pseudo-autosomal region (PAR) of the short arm of the X-chromosome. Only the first exon encompasses the non-PAR region, while the rest of the gene is located in the shared region between the X and Y chromosome³⁷⁷. In contrast, the human *MID1* gene is located in the Xp22.2 region, near but not within the PAR region^{368,378,379}. Both human and murine genes consist of 10 exons, with 9 coding exons. Additionally, there are at least 5 alternative promoters in the 5' end, leading to alternative transcription of untranslated exons³⁸⁰. *MID1* isoforms exhibit differences in the 5' region, the polyadenylation signal, and the coding region^{380–382}. *MID1* transcription is epigenetically regulated by several miRNAs, as identified in studies conducted in HEK193 cells and lymphoblastoid cells, including miR-19, miR-340, miR-374, miR-542, miR135b^{383,384}.

1.5.1 Protein function

The human MID1 protein is translated from a large mRNA molecule of more than 6400 nucleotide, resulting in a protein of 667 amino acid residues. As part of the TRIM family, the MID1 protein comprises three characteristic domains in its N-terminal region: a RING domain, two B-boxes domains, and a coiled-coil region (CC)³⁸⁵. Following the CC domain, the protein possesses a COS domain (C-terminal subgroup one signature), Fibronectin type III repeat, and in the C-terminal a PRY domain associated with the SPRY domain^{368,385,386}.

The MID1 protein serves several functions within cells, including acting as mRNA binding regulating translation, E3 ubiquitin ligase, and associating with microtubules³⁸⁷. The latter function was discovered in 1999 and involves the CC and COS domain^{386,388–390}. The association of the MID1 protein with microtubules allows its transport through dynein or kinesin proteins. Phosphorylation of MID1 amino acid residues can alter its function, including its binding to microtubules. Studies have shown contradictory results regarding the role of MID1 phosphorylation on microtubule association, with some indicating that it favours binding to microtubules³⁹¹, while others suggest it inhibits it³⁹². These discrepancies may be explained due to different MID1 phosphorylation sites investigated or experimental conditions. The exact role of the microtubule association of MID1 remains unclear. Some studies indicate that when pathogenic mutations impair MID1 binding to microtubules, microtubule organization is unaffected^{388,389}, while others suggest that MID1 has a protective role against microtubule depolymerisation when ectopically expressed³⁸⁹. This protective role appears to be associated with the partner protein MIG12 (MID1 interacting G-12 like protein, also known as MID1IP1). When co-transfected and ectopically expressed in cells, MIG12 is found associated with

microtubules alongside MID1, and their synergistic role is associated with depolymerization protection³⁹³. Since many of these studies relied on overexpressed systems, it is hypothesized that in physiological systems, the association and microtubule protection may be a regulated mechanism influenced by phosphorylation and/or other proteins' expression. Furthermore, the association of MID1 with microtubules is maintained throughout the cell cycle and cell division^{388,394-396}. Studies have shown that the absence of MID1 leads to cytokinetic defects³⁹⁴, suggesting a role for MID1 in regulating microtubule dynamics during cell division and thus controlling processes such as cell adhesion, cell migration, and symmetric and asymmetric cell division^{397,398}.

Post-translational ubiquitination is a multi-faceted modification involved in various functions, including the well-known protein degradation via the proteasome-dependent pathway, as well as the modification of protein localization, activity, and interactions³⁹⁹⁻⁴⁰¹. Protein ubiquitination generally occurs on lysine residues, where the ubiquitin protein is enzymatically linked to the amino acid, resulting in mono- or poly-ubiquitination that can mediate different roles⁴⁰². Ubiquitin, a small protein, contains several lysine residues that can be ubiquitinated (K6, K11, K27, K29, K33, K48, K63), allowing for the creation of complex chains with various morphologies and ramifications that translate into specific functions. For example, K63 ubiquitination is associated with NF- κ B signalling, endocytosis, and DNA repair, while K48 ubiquitination is associated with classical proteasome-mediated protein degradation⁴⁰². The mechanism by which a protein is ubiquitinated involves three enzymes: E1 enzymes, responsible for ubiquitin activation; E2, also known as ubiquitin-conjugating enzymes; and E3 ubiquitin ligases, which bind to specific protein requiring ubiquitination⁴⁰³. E3 enzymes are categorized based on their domains that exhibit E3 activity, either the RING or the HECT domains⁴⁰⁴⁻⁴⁰⁶. As part of the TRIM family, the MID1 protein has a RING domain, with RING-dependent autoubiquitination activity empirically demonstrated in 2011⁴⁰⁷⁻⁴⁰⁹. Its ubiquitination activity is essential for interacting with different E2 proteins thereby functionally regulating several proteins.

In 2001, Trockenbacher and colleagues discovered that MID1 interacts with a protein known as alpha-4, a non-canonical part of PP2A (protein phosphatase 2A)^{410,411}, revealing that this interaction is mediated by the first B-box domain of MID1 and the C-terminal domain of alpha-4^{370,391}. Studies conducted on fibroblasts derived from XLOS patient demonstrated that the presence of MID1 mutations resulted in elevated levels of PP2Ac, the catalytic subunit of PP2A, leading to an increase of its phosphatase activity, particularly on microtubule-associated proteins. Conversely, ectopic expression of MID1 led to a reduction in PP2Ac levels. These findings suggest a link between ubiquitination, the recruitment of alpha-4 to MID1-associated microtubules, and the formation of a MID1/alpha-4/PP2A complex, ultimately resulting in the proteasome-mediated degradation of PP2Ac via polyubiquitination³⁷⁰. However, alpha-4 also possesses a ubiquitin interacting motif (UIM) that, when monoubiquitinated, confers a protective effect against the polyubiquitination of PP2Ac^{411,412}. It has been demonstrated that MID1 exploits this protective mechanism by monoubiquitinating alpha-4 on its C-terminal site, leading to a cleavage of alpha-4 by calpain. This cleavage seems to be necessary for the poly-ubiquitination and subsequent degradation of PP2Ac^{408,413}.

1.5.2 MID1 regulates the mTOR pathway

The MID1/alpha4/PP2A complex plays a crucial role in regulating the mTOR (mammalian target of rapamycin) pathway, which is involved in various cellular processes including autophagy, protein synthesis, cell metabolism, growth, proliferation, intracellular transport, and cell migration⁴¹⁴⁻⁴¹⁶. In XLOS patients, the absence of MID1 results in elevated PP2A levels, leading to the dephosphorylation of mTOR protein, disruption of the mTORC1 complex, decreased phosphorylation of S6K1, and impaired cell growth and protein translation. This phenotype can be recapitulated by MID1 depletion *in vitro*, and can be rescued by re-introducing wild type MID1⁴¹⁵.

Another sets of MID1 downstream targets comprises RNA molecules, which are bound by the MID1/alpha4 complex through MID1-association sequences (MIDAS), which are RNA- hairpin structures containing purine-rich sequences^{392,417,418}. Once bound by the MID1 complex, these target RNA molecules are positioned closer to ribosomes, and their translation is facilitated^{383,392,417,419,420}. Among the various RNAs targeted by the complex, specifically *PDPK-1* RNA is regulated by MID1, thereby controlling the mTOR pathway. Since the mTOR signalling pathway is regulated in part through *PDPK-1* translation, MID1 complex-mediated enhancement of *PDPK-1* influences several processes such as mTOR⁴²¹. Studies on systems lacking MID1 have confirmed reduced levels of *PDPK-1* translation, which can be restored to normal levels upon reintroduction of MID1 into the system⁴¹⁸.

The mTOR pathway is also implicated in the regulation of the Sonic Hedgehog (Shh) pathway. In *Drosophila*, hedgehog represents a group of signalling molecules: Sonic Hedgehog (Shh), Indian hedgehog (Ihh), and Desert hedgehog (Dhh). These molecules play crucial roles in regulating signalling pathways essential for cellular proliferation and differentiation, particularly during embryonic development, where they contribute to the formation of the body axis by establishing concentration gradients⁴²². Upon secretion, these molecules bind to the transmembrane protein Patched (Ptch), initiating a signalling cascade within the target cell. Ptch activation leads to the activation of another transmembrane protein, smoothened (Smo), which transduces the signal intracellularly, thereby modifying the activity of protein such as KIF7 and SuFu (suppressor of fused). This cascade ultimately results in the nuclear translocation of Gli transcription factors⁴²³, facilitating the translation of the signal into gene expression. Studies in human cells have demonstrated that the reduction of mTOR activity, achieved through the use of rapamycin (a drug that inhibits the mTORC1 complex), can diminish the nuclear translocation of GLI3. Consequently, this reduces the expression target genes regulated by Gli transcription factors, such as Cyclin D1⁴²⁴. The MID1 RNA/protein complex plays a crucial role in regulating GLI3 activity. Inhibition of MID1 activity, resulting in elevated levels of active PP2A phosphatase, leads to increased cytoplasmic concentration of GLI3 compared to the nucleus. Conversely, overexpression of the MID1-interacting protein alpha4 promotes the translocation of GLI3 into the nucleus^{424,425}. The link between MID1 complex activity and GLI3 nuclear translocation was elucidated in 2014, when it was also demonstrated that MID1 binds to and ubiquitinates Fu protein, a component of the Shh pathway in mammalian cells. This ubiquitination leads to the proteasome-dependent cleavage of Fu protein, generating two fragments. The N-terminal Fu fragment enhances GLI3 nuclear localization and augments Shh-dependent gene expression⁴²⁶. The interaction between MID1 and the Shh pathway is particularly significant during embryonic development, where these molecules regulate the ventral midline of the embryo. Disruptions in this process have been observed in XLOS patients and in studies of *Xenopus* and chicken development. In *Xenopus*, Shh expression induces the upregulation of *MIDI1* in the forebrain and optic cup⁴²⁷. Similarly, in chickens, there exists a feedback loop that regulates the expression of *MIDI1* and SHH, where low levels of SHH repress *MIDI1* transcription, while moderate levels of SHH enhance *MIDI1* expression⁴²⁸.

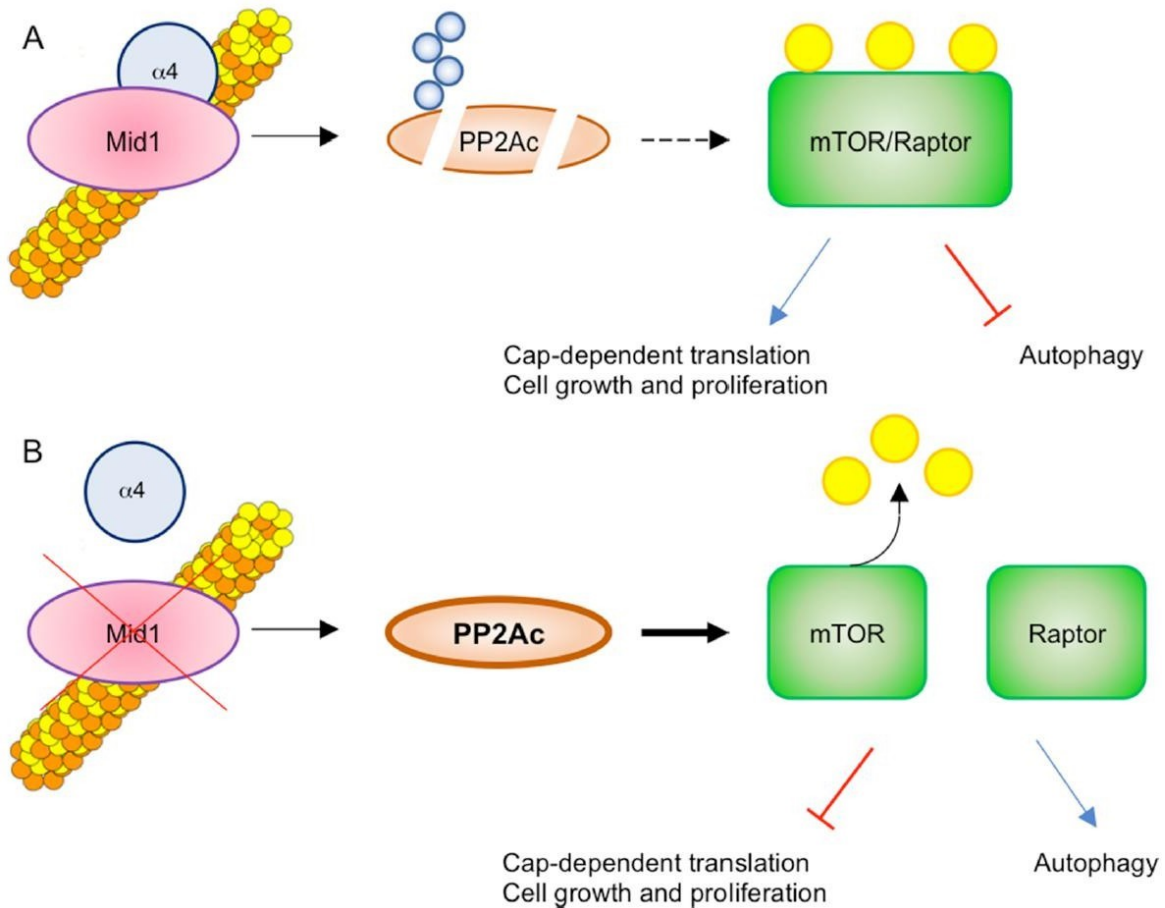


Figure 8. mTOR pathway is regulated by MID1 activity. Figure adapted from Baldini, Mascaro, and Meroni, 2020³⁸⁷. Cartoon representing the role of MID1 in regulating mTOR pathway. When PP2Ac is degraded by MID1 activity, association between mTOR and Raptor can take place, leading to the activation of mTOR signalling pathways. Loss-of-function of MID1 increase PP2Ac protein levels, inhibiting the mTORCH1 complex formation.

1.5.3 Mutations affecting *MID1* gene

The *MID1* gene is associated with XLOS disease. Studies of mutations in patients have revealed a complex and diverse array of mutations within the *MID1* gene, including missense and nonsense mutations, mutation affecting splicing sites, indels, and large deletions. Some patients have been reported to carry complete or partial deletion of the *MID1* gene. Notably, the majority of mutations affect the C-terminal part of the protein, indicating a distinct impact of mutations affecting the RING domain^{368,372,376,429–431}. The various types and locations of *MID1* mutations suggest that loss of function of MID1 protein contributes to the phenotype in XLOS disease in patients.

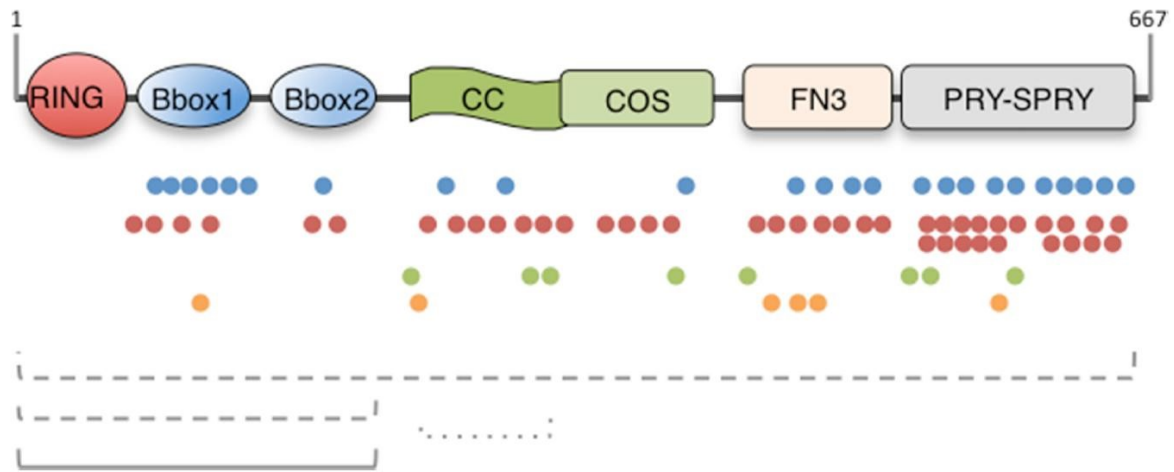


Figure 9. MID1 protein structure and reported mutations associated with XLOS disease. Figure adapted from Baldini, Mascaro, and Meroni, 2020³⁸⁷. The 667 amino acid sequence of MID1 protein comprises the RING domain (10-59 amino acids), B-Box1 (114-164 amino acids) and B-Box2 (170-212 amino acids), CC (219-319 amino acids), COS (320-380 amino acids), FN3 (382-472 amino acids), and PRY (483-528) linked to SPRY (538-657). Mutations are reported as blue dots for missense mutations, red dots for nonsense and truncating mutations, green dots for splice site mutations, orange dots for in frame indels, and dashed lines for deletions and rearrangements, while continuous line for duplication.

2 Aim of the thesis

Neurodevelopmental Disorders (NDDs) exhibit a strong sex bias in human, with male being more frequently and more severely affected than female⁴³². Although theories were postulated to explain the sex bias observed in NDDs, the mechanism underlying male vulnerability or female resilience is still not known⁴³³⁻⁴³⁹. The main factors that contribute to the sex differences in NDDs are the circulating sex hormones and the differences in the sex chromosome complement between female (XX) and male (XY)⁴⁴⁰. Recent studies shed light on the role of sex hormones in the developing brain⁴⁴¹, while the enrichment of NDD-associated genes on the X-chromosome compared to the autosomes suggests a key role of the X-chromosome in the sex differences of NDD⁴⁴².

X-chromosome inactivation (XCI) is responsible for the gene dosage compensation of X-linked genes between females (XX) and males (XY) by transcriptionally silencing one of the two X-chromosomes in females⁴⁴³⁻⁴⁴⁶. However, the silencing is not complete and more than 20% of X-linked genes are constantly or variably expressed from the inactive X-chromosome⁴⁴⁷. Among tissue-specific escape genes, genes eluding the XCI in the brain and during neurogenesis were identified in mouse^{448,449}. However, no studies shed light into the facultative escape gene dynamic during human neurodevelopment.

Preliminary data from my working group (Thesis, Dr. Stephan Kaseberg) showed that the X-linked *MIDI* gene is re-expressed from the inactive X-chromosome during in vitro 2D neural differentiation. To study the neurodevelopmental impairment caused by mutation in the *MIDI* gene, iPS clone derived from a patient heterozygous for a *MIDI* mutation were generated and differentiated into Neural Progenitor Cells (NPCs) and neurons. Surprisingly, the expression of the *MIDI* allele located on the inactive X-chromosome was detected after in vitro neural differentiation.

Based on previous identification of *MIDI* as a human facultative escape gene during neural development, in my PhD thesis I investigated the presence of a dynamic gene expression from the inactive X-chromosome during neural development, potentially leading to changes of the gene expression pool in females, thus allowing higher flexibility of the regulation of protective genes against Neurodevelopmental Disorders.

To investigate the dynamic expression of the inactive X-chromosomal genes during neural differentiation, three iPSC lines, derived from three unrelated female donors and showing a clonal status with skewed XCI, were differentiated by following a previously established 2D neural differentiation protocol. The transcriptome of the iPSCs, NPCs, and neurons was compared, focusing the analysis on the allele-specific expression of X-linked genes and identifying different categories of facultative escape genes. Further bioinformatic analyses were performed to investigate the position of the identified escape genes along the X-chromosome, their relation with the NDDs and previously identified escape genes, and their impact on the whole genome.

Furthermore, validation of the RNA-sequencing results was performed by employing allele-specific RT-PCR, QUASEP assay, and single-cell resolution RNA-FISH imaging for the visualization of the allele-specific expression of facultative escape genes. The physiological relevance of the results obtained *in vitro* was strengthened by tackling one of the most common caveats of using hiPSCs in vitro system, the erosion of the X-chromosome.

3 Material and Methods

3.1 Material

3.1.1 Equipment

Table 1 Laboratory equipment used during the PhD thesis

Name	Manufacturer	Model	Function
Nanodrop™ OneC Spectralphotometer	ThermoFisher Scientific	ND-ONEC-W	DNA and RNA quantification
StepOnePlus™ Real-Time PCR System	ThermoFisher Scientific	4376600	Real time qPCR
PeqPower 300 V	Peqlab	300 V	Powersupply for agarose chamber
Electrophorese chamber	Peqlab	40-1214, 0-150V, 0-100mA ClassII	Agarose gel separation
Gel iX20 Imager	Intas Science Imaging	iX20	UV imaging of agarose gel
Agarose gel comb	VWR / PeqLab	20 wells 1.5 mm	Preparation of agarose gel
PyroMark Q96 instrument	Qiagen	9001525	Pyrosequencing
NextSeq	Illumina	NextSeq500	RNA-sequencing
HybEZ™ II Oven	ACD	II	RNA-FISH
Visiscope 5 elements	Visitron System GmbH	5 elements spinning disk confocal	RNA-FISH imaging
Ti-2E stand	Nikon	Ti-2E	RNA-FISH imaging
sCMOS camera	Photometrics	BSI	RNA-FISH imaging
BC43 spinning disk confocal	Oxford Instruments	BC43	RNA-FISH imaging
Revolve Microscope	Echo	Revolve	Microscope
Olympus SZX10 Stereo Microscope	Olympus	SZX10	Stereomicroscope
Leica BGV A3 Stereo Microscope (2011)	Leica	BGV A3	Stereomicroscope
Axiovert25 Inverted Microscope	Zeiss	Axiovert25	Inverted microscope
EVOS XL Microscope with photography	Life Technologies	EVOS XL	Inverted microscope
CO2 incubator	Binder		Cell incubator
CO2 incubator	ThermoForma	371	Cell incubator
Revco EXF	Binder Thermo Scientific	EXF	Cell incubator
ARPEGE140 Liquidi nitrogen tank, LIQUIDE, Air Liquide	Arpege	ARPEGE 140	Cell vial long-term storage
Neubauer chamber	Brand		Cell counting
4D-Nucleofector	Lonza	AAF-1002B+AAF-1002X	iPSC electroporation

BD FACS Aria TM III Cell Sorter	BD	FACS Aria TM III	Cell sorter for electroporated cells
ACD HybEZ II Hybridization System (220v) with ACD EZ-Batch Slide System	Bio-Techne GmbH	II Hybridization System	Oven for RNA-FISH procedure
Liebherr Fridge and Freezer Combination & Freezer	Liebherr		Samples and reagents storage
RevcoExF	Thermo Fisher Scientific	8930	
Acco-Jet Pro	Brand	Pro	Pipette boy
ErgoOne Fast	Starlab	FAST	Pipette boy
Rollermixer SRT9D	Stuart	SRT9D	Roller for mixing solutions
HERASafe Heraeus Sterile Working Bench	Thermo Fisher Scientific	HERASafe	Sterile cell culture working bench
Advanced Primus 96 Thermocycler	PeqLab	96-well	Thermocycler
T100 TM Thermal Cycler	BioRad	T100	Thermocycler
VV3	VWR	VV3	Vortexer
Vortex Genie 2 TM	Bender & Hobein AG	Genie 2	Vortexer
Scout TM Pro 600g	Ohaus	DB-4726.5	Scales
Mini Star Table Centrifuge	MiniStar	Mini Star	Table centrifuge
Roth Table centrifuge	Roth	ROTILABO Uni-fuge	Table centrifuge
Perfect Spin P Centrifuge	PeqLab	PerfectSpin P	Table centrifuge for 96-well plates
Eppendorf centrifuge 5415D	Eppendorf	5415D	Centrifuge
Eppendorf centrifuge 5415C	Eppendorf	5415C	Centrifuge
Perfect Spin 24R Refrigerated microcentrifuge	PeqLab	Perfect Spin 24R	Centrifuge
Heraeus Megafuge 16R	Thermo Scientific	Megafuge 16R	Centrifuge

3.1.2 Disposable material

Table 2 Laboratory disposables used during the PhD thesis

Name	Manufacturer	Description
12-well cell culture plate Cell star	Greiner Bio-One	12 well plate
6-well cell culture plate Cell star	Greiner Bio-One	6 well plate
Safe-Seal Micro tube (1.5 ml)	Sarstedt	reaction tube,
Eppendorf microfuge tubes (1,5/2 ml)	Eppendorf	Reaction tube
Micro screw tube	Sarstedt	Tubes for aliquoting media supplements (e.g. N2, FGF, ...)

Cell Star tubes (15/50 ml) (PP, graduated, conical bottom, blue screw cap, sterile)	Greiner Bio-One	15 ml and 50 ml falcons
0.2 ml 8-Strip Non-flex PCR Tubes, with Flat Caps	StarLab	PCR tubes
Cryo vials (Self-sealing Cap)	StarLab	Cryo preservation / storage of frozen cells
TipOne filter Tips (10/20/200/1000 µl)	TipOne	Pipette tips
StarLab Filter Tips (10/20/200/1000 µl)	StarLab	Pipette tips
(Nonfiltered) Yellow Tips (200 µl)	StarLab	Pipette tips used in cell culture for removing media with glass pipettes and pump
SafeSeal SurPhob Tips, Low Binding (10 µl)	Biozym	Pipette tips for RT-qPCR
Rack for tubes (1.5/2ml)	NeoLab	Racks for microfuge tubes
NeoRack for falcons	Neolab	Racks for 15 ml and 50 ml falcons
Discovery Comfort Pipettes (10/20/200/1000 µl)	Discovery Comfort	Pipettes
Ergo One Pipettes (10/20/200/1000 µl)	StarLab	Pipettes
Eppendorf Research Pipettes (10/20/200/1000 µl)	Eppendorf	Pipettes
Pipettes (5ml, 10ml, 25ml with graduation)	Greiner Bio-One	glass pipettes
Steri Cup Quick Release, Millipore Express Plus 0,22 µM (Filter Bottles)	Merck Millipore	Bottles to sterile filter media
Superfrost Microscope slides	Thermo Scientific	Glass slides for immune florescence staining
Parafil™M™	Bernis	Laboratory film
Nitril gloves	StarGuard	Sterile gloves
PCR adhesive seal sheet	Thermo Scientific	Sealing of 96-well plates for RT-qPCR
Cell Culture Dish Cell Star	Greiner Bio-One	Petri dishes in cell culture
175cm ² Cell Culture Flasks	Greiner Bio-One	Fibroblast culturing
TC Flask T75, Cell+	Sarstedt	Fibroblast culturing
Microtube TOUGH-TAGS	Diversified Biotech	Tube labelling
Microtube TOUGH-SPOTS	Diversified Biotech	Tube labelling
Multipette Plus manual handdispenser	Eppendorf	Automatic dispenser for QUASEP assay
Multichannel pipette (200 µl)	Thermo Fisher	Multichannel dispenser for PCR products during QUASEP assay

3.1.3 Chemicals and Media

3.1.3.1 Chemicals

Table 3 Chemicals used during the research project

Name	Manufacturer
Acetic Acid	
Agarose	AppliChem
Ampuwa Spüllösung, 1000ml Plastipur	FRESENIUS KABI
Aqua ad iniectabilia Braun	BRAUN
Boric acid	Carl Roth
BSA (Bovine serum albumin)	Carl Roth
CaCl ₂	Carl Roth
DPBS	Gibco LifeTechnologies
EDTA Disodium Salt 2-hydrate	AppliChem
99,99% Ethanol	Honeywell, Riedl-de Haen
99% Ethanol (for disinfection 70 % in water)	Martin & Werner Mundo oHG
Ethidiumbromide	Carl Roth
HEPES	Carl Roth
Isopropanol	Carl Roth
Meliseptol Foam Pure	Carl Roth (Braun)
NaCl	Carl Roth
NaOH	Carl Roth
PFA	Carl Roth
Poly-L-Ornithine-hydrobromide	SIGMA-Aldrich
ProLong™ Diamong	Thermo Fisher
ProLong™ Gold	Thermo Fisher
ProLong™ Glass Antifade	Thermo Fisher
Proteinase K	NEW ENGLAND BioLabs & AppliChem
SDS	Carl Roth
Streptavidin-Sepharose beads	GE Healthcare
RNase AWAY (# 7002)	M&B Molecular Bio Products
Terralin liquid	Schülke +
TRIS	Carl Roth
Triton-X 100	Carl Roth
Vectashield	Vector Laboratories
PyroMark Gold Q96 Enzyme	QIAGEN
PyroMark Gold Q96 Substrate	QIAGEN
PyroMark Gold Q96 dNTPs (dATP, dGTP, dCTP, dTTP)	QIAGEN
Binding buffer	(Generated in the diagnostics laboratory)
Annealing buffer	(Generated in the diagnostics laboratory)

3.1.3.2 Media

Table 4 Media and compounds used for cell culturing during the PhD project

Name	Manufacturer
Accutase	ThermoFisher Scientific
Advanced DMEM	Gibco LifeTechnologies
B27+VitA-supplement	Gibco LifeTechnologies
B27-supplement	Gibco LifeTechnologies
bFGF	Gibco LifeTechnologies
CloneR™ 10x Cloning Supplement	Stemcell Technologies
DMEM (1x) high glucose	Gibco LifeTechnologies
DMEM	Gibco LifeTechnologies
DMEM with GlutaMAX™	Gibco LifeTechnologies
DMEM/F-12 GlutaMAX™	Gibco LifeTechnologies
DMSO	Carl Roth
DPBS	Gibco LifeTechnologies
ESC-qualified FBS	Gibco LifeTechnologies
FBS	Gibco LifeTechnologies
Gelatine	Sigma
Geltrex	Gibco LifeTechnologies
HBSS	SIGMA
IMDM	Gibco LifeTechnologies
KOSR	Gibco LifeTechnologies
Laminin	Sigma
Matrigel Matrix	Corning
mTeSR™1	Stemcell Technologies
mTeSR™1- supplement	Stemcell Technologies
N2-supplement	Gibco LifeTechnologies
MEM-NEAA	Gibco LifeTechnologies
Neural Induction Supplement	Gibco LifeTechnologies
Neurobasal Medium	Gibco LifeTechnologies
Opti-MEM	Gibco LifeTechnologies
Pen/Strep	Gibco LifeTechnologies
Poly-L-Ornithine-hydrobromide	SIGMA
Rock Inhibitor	Stemcell Technologies
RPMI 1640	Gibco LifeTechnologies
TrypLETMExpress	Gibco LifeTechnologies

3.1.3.3 Media composition

Table 5 Media composition and function for different cell types used during the PhD thesis

Cell type	Media function	Media composition
Fibroblasts	Extraction	DMEM + 20% FBS + 1% Pen/strep
	Culturing	IMDM +1% Pen/Strep + 15% Fetal bovine serum (FBS)
	Freezing	Culturing Medium + 10% DMSO
	Fibroblast	FRM: Fibroblast Reprogramming Medium:

iPSC reprogramming		DMEM + 10% ESC-Qualified FBS + 1% MEM Non-Essential Amino Acids Solution (10 mM) + 0,1% β -mercaptoethanol (55 mM)
	MEF cells	MEF medium = FRM
	iPSCs	iPSC Reprogramming Medium: DEMEM/F-12 (with GlutaMAX) + 20% KnockOut Serum Replacement (KOSR) + 1% MEM Non-Essential Amino Acids Solution (10 mM) + 0,1% β -mercaptoethanol (55 mM) + 1% Pen/Strep + 0,4% bFGF (10 μ g/ml)
iPSCs	Coating	Matrigel Matrix or geltrex diluted in DMEM/F-12 (with GlutaMAX)
	Replating	PBS/EDTA: 500 ml DPBS + 301 μ l 0,8M EDTA solution 0,9g NaCl
	Culturing	mTeSR _{TM} 1 + mTeSR _{TM} 1 supplement + 1% Pen/Strep
	Freezing	iPSC culturing medium + 10 % DMSO + 25% KOSR
NPCs differentiation	Neural induction	NIM: Neural Induction Medium Neurobasal medium + 2% Neural induction supplement + 1% Pen/strep
	Neural expansion	NEM: Neural Expansion Medium: ~50 % Neurobasal medium + ~ 50% Advanced DMEM/F-12 + 2% Neural induction supplement + 1% Pen/strep
NPCs	Coating	Poly-Ornithine/Laminin
	Culturing	NM: Neural Medium DMEM/F-12 + 1% Pen/Strep + 2% Vitamin B27 + 1% N2-solution + 0,1% FGF2
	Replating	NM + 15-30% KOSR
	Freezing	NM + 10% KOSR + 10% DMSO
Neurons	Coating	Poly-Ornithine/Laminin
Neurons	Culturing	Neurobasa +: Neurobasal Medium + 1% Pen/Strep + 1% Glutamax + 2% Vitamine B27 with VitA

3.1.4 Reagents, Kits, and Enzymes

3.1.4.1 Kits and reagents

Table 6 Reagents and kits used during the PhD thesis

Name	Manufacturer
CytoTune TM -iPS 2.0 Sendai Reprogramming Kit	Thermofisher Scientific
High Pure RNA Isolation Kit	Roche
PrimeScript TM RT Master Mix	TaKaRaBio
Revert Aid First Strand cDNA Synthesis Kit	Thermofisher Scientific
PyroMark Gold Q96 Reagents	QIAGEN

Reverse Transcription Kit with PrimeScript™ RT Master Mix	TaKaRaBio
Ovation® SoLo RNA-Seq Library Preparation Kit	NuGEN
P3 Primary Cell 4D-Nucleofector™ X Kit L	Lonza
PSC Neural Induction Kit	Thermofisher Scientific
RNA Cleanup Kit	Monarch
RNAscope Multiplex Fluorescent Detect Kit V2	Bio-Techne
RNAscope wash buffer reagents	Bio-Techne
RNAscope H202 and Protease Reagents	Bio-Techne
TSA buffer	Bio-Techne
Probe Diluent	Bio-Techne
Opal Dye 520	Akoya
Opal Dye 570	Akoya
Opal Dye 690	Akoya

3.1.4.2 Enzymes

Table 7 Enzymes used during the PhD thesis

Name	Manufacturer
FastStart™ Taq DNA Polymerase	Roche
PrimeScript™ RT Master Mix	TaKaRaBio
SYBR® Premix Ex Taq™ II	TaKaRaBio
Proteinase K	NEW ENGLAND BioLabs & AppliChem
Enzyme	PyroMark Gold Q96 Reagents, QIAGEN
RNase A, DNase and protease free (10 mg/ml)	Thermo Fisher

3.1.5 Primers and Antibody

3.1.5.1 Primers

Table 8 Primers used during the PhD thesis. Primers were ordered with a concentration of 100 µM and diluted with dH₂O to 10 µM before used.

Name	Sequence (5' → 3')
Allele-specific RT-PCR	
MID1_ Ex8_ WT_ for	AGTGGCCGGCATTATTGGGAAGTGGT
MID1_ Ex9_ WT_ rev	ATGCCACGCGCCGGAGGT
MID1_ Ex9_ MUT_ rev	AGGATGCCACGCGCCGT
RT-qPCR	
GAPDH_ for	CCACATCGCTCAGACACCAT
GAPDH_ rev	AAATCCGTTGACTCCGACCTT
Klf4_ for	CCCACATGAAGCGACTTCCC
Klf4_ rev	CAGGTCCAGGAGATCGTTGAA
MAP2_ For	GGAGGTGTCTGCAAGGATAGT
MAP2_ Rev	GGTGGAGAAGGAGGCAGATT
NANOG_ for	AAGGTCCCGGTCAAGAAACAG
NANOG_ rev	CTTCTGCGTCACACCATTGC
Nestin_ for	CCAGATCGCTCAGGTCTG

Nestin_rev	AGCTGAGGGAAGTCTTGGAG
Oct4_for	GTGTTTCAGCCAAAAGACCATCT
Oct4_rev	GGCCTGCATGAGGGTTTCT
PAX6_for	ACCCAAGAGCAAATTGAGGC
PAX6_rev	CCATTTGGCCCTTCGATTAGA
SOX2_for	TGGACAGTTACGCGCACAT
SOX2_rev	CGAGTAGGACATGCTGTAGGT
TAU_For	GTGCAAATAGTCTACAAACCAGT
TAU_Rev	CAATCTTCGACTGGACTCTGT
TUBB_For	TCGGACTTGCAGCTGGAG
TUBB_Rev	CAGGCCTGAAGAGATGTCCA
Xist For	GTAGGTGTGCTGATAACCAAGGC
Xist Rev	GGGAAAGGAAGATTGAGGGTGG
XACT_1_For	GTGAGTGTTTCATGGATTAGGGC
XACT_1_Rev	TTGGCTTCAGCCCTCATTGT
XACT_2_For	TGAGTGTTTCATGGATTAGGGCAT
XACT_2_Rev	TGGCTTCAGCCCTCATTGTT
PCR for QUASEP	
MID1_For	AATAACTGGGTGGTGAGACACA
MID1_BioRev	[Btm]AGGCGATAGAGCCGTTAT
MID1_AQ_BioFor	[Btm]CAGGAAAATGGGCCAGAC
MID1_AQ_Rev	CGTAGTAATTCCTGAGGCAGAAA
GPM6B_micro_BioFor	[Btm]AAAATTGCTTCTGGAAAGCTTGTC
GPM6B_micro_Rev	CCCAAGTCCAGGACCCATTAATT
GPM6B_RNA-seq_For	TGGCACTGATAGAAAATATTGATT
GPM6B_RNA-seq_BioRev	[Btm]TCCTGTTGCATGACAGATTTAACT
ZNF185_For	AGGCTTATAATGGGCCAAGTTGA
ZNF185_BioRev	[Btm]CAACACATGCCAACATACCTGTAA
QUASEP	
MID_Seq	GAGCCTGCCCCCAC
GPM6B_micro_S	ATCTGGATGCTGGACT
GPM6B_RNA-Seq_S	AGCTACCAGAATTAGAAAAG
ZNF185_Seq	AGTTTTTCATTTGGTCTTC
MID1_AQ_Seq	AAAAGATAATTCCAGGAAG
SeV vectors clearance RT-PCR	
SeV_For	GGATCACTAGGTGATATCGAGC
SeV_Rev	ACCAGACAAGAGTTTAAGAGATATGTATC
KOS_For	ATGCACCGCTACGACGTGAGCGC
KOS_Rev	ACCTTGACAATCCTGATGTGG
KLF4_For	TTCCTGCATGCCAGAGGAGCCC
KLF4_Rev	AATGTATCGAAGGTGCTCAA
c-MYC_For	TAACTGACTAGCAGGCTTGTCG
c-MYC_Rev	TCCACATACAGTCCTGGATGATGATG

3.1.5.2 Primary antibodies

Table 9 Primary antibodies used during the PhD project

Protein	Organism	Source	Identifier
α -SERPINH1	Rabbit	Sigma-Aldrich	S5950
α -NANOG	Goat	R&D Systems	AF1997
α -TRA-1-60	Mouse	Millipore	4360
α -H3K27me3	Rabbit	Cell Signaling Technologie	9733
α -PAX6	Rabbit	BioLegend	901301
α -SOX2	Rabbit	Abcam	ab137385
α -NESTIN	Mouse	Merck	MAB5326
α -MAP2	Mouse	Sigma	M4403
α -TAU	Mouse	Abcam	ab32057
α -TUBB3	Rabbit	Sigma Aldrich	T8660

3.1.5.3 Secondary antibodies

Table 10 Secondary antibodies used during the PhD project

Protein	Organism	Source	Identifier
anti-Rabbit IgG, Alexa Fluor 488	Goat	Invitrogen	A11008
anti-Mouse IgG, Alexa Fluor 488	Goat	Invitrogen	A11017
anti-Mouse IgG, Alexa Fluor 488	Rabbit	Invitrogen	A11008
anti-Mouse IgG, Alexa Fluor 594	Goat	Invitrogen	A11012
anti-Goat IgG, Alexa Fluor 594	Rabbit	Invitrogen	A11079

3.1.5.4 RNA-FISH probes

Table 11 RNA-FISH probes used during the PhD project.

Gene	Organism	Nomenclature	Identifier
GPM6B	Human	Hs-GPM6B-intron-C3	1201031-C3
MID1	Human	Hs-MID1-O1-C3	1224591-C3
MID1	Human	Hs-MID1-O1-C2	1224591-C2
NLGN4X	Human	Hs-NLGN4X-O1-C3	1180201-C3
SPIN3	Human	Hs-SPIN3-Intron-C3	1329651-C3
XACT	Human	Hs-XACT-C1	471031-C1
XIST	Human	Hs-XIST-C2	311231-C2

3.1.6 Plasmids and gRNA constructs

Table 12 Plasmids used during the PhD project

Name	Manufacturer/Website
pU6-(BbsI)sgRNA_CAG-925 Cas9-venus-bpA	Addgene #86986
gRNA_Cloning Vector	Addgene #41824

Table 13 gRNAs used during the PhD project. The black sequences are referred to the T7 promoter sequence necessary for cloning the gRNAs in the gRNA plasmid. The red and violet sequences are the gRNA specific sequence for targeting the DNA. * referred to phosphorothioate linkages.

Name	Sequence
MID1_Ex9_gR NA1_For	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCCGATGCC ACGCGCCGGAGGT
MID1_Ex9_gR NA1_Rev	GACTAGCCTTATTTAACTTGCTATTTCTAGCTCTAAAACACCTCCGG CGCGTGGGCATC
MID1_Ex9_gR NA2_For	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCCGGATGCC CACGCGCCGGAGG
MID1_Ex9_gR NA2_Rev	GACTAGCCTTATTTAACTTGCTATTTCTAGCTCTAAAACCTCCGGC GCGTGGGCATCC
MID1_delCTCC _deletion_templ ate	A*A*AAGGCGATAGAGCCGTTATCATAGTCCAGCAGGATGCCACGC GCCGTGGGGGCAGGCTCAATGGGGATTTCCTTGCTATTGTGTCTCA CC*A*C

3.1.7 Softwares and online tools

Table 14 Software and online tools used during the PhD project

Name	Manufacturer/Website
Bioedit	mBio
Office	Microsoft
Prism	GraphPad
Pyro Q CpG	Qiagen
Intas gDS	Intas Science Imaging
ND_1000 3.5.1	Peqlab Biotechnologies
NCBI	www.ncbi.nlm.nih.gov
Genome Browser (USCS In-Silico PCR)	www.genome.ucsc.edu
Varsome	www.varsome.com (hg19)
Primer-Blast	https://www.ncbi.nlm.nih.gov/tools/primer-blast/
Ensembl	http://www.ensembl.org/index.html https://grch37.ensembl.org/index.html /
Refseq	https://www.ncbi.nlm.nih.gov/refseq/
VisiView software	Visitron System GmbH
Fiji - ImageJ-win64	https://imagej.net/software/fiji/
Imaris Image Software	Oxford Instruments
Inkscape	Sodipodi
Adobe Illustrator	Adobe

bcl2fastq conversion software	Illumina
Cutadapt	https://cutadapt.readthedocs.io/en/stable/
FastQC	
STAR aligner	
nudup.py	NuGen
BBDuk	BBMap
featureCounts	https://rnh.github.io/bioinfo-notebook/docs/featureCounts.html
Clara Parabricks Pipeline	NVIDIA
Haplotype Caller	https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller
Variant Effect Predictor	Ensembl
VariantsToTable	https://gatk.broadinstitute.org/hc/en-us/articles/360036896892-VariantsToTable
Meta R-package	https://cran.r-project.org/web/packages/meta/index.html
karyoploteR package	https://bernatgel.github.io/karyoploteR_tutorial/
igraph package	https://r.igraph.org/
clusterProfiler package	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
ChromDiff tool	https://github.com/angieyen/ChromDiff
DOSE R package	https://www.bioconductor.org/packages/release/bioc/html/DOSE.html

3.1.8 Cells

Table 15 Cell lines used during the PhD project

Patient/Donor	Donor age	Sex	Cells	Name	Cell type	mutation
16/98	~30 years	Female	16/98	Fibro	Fibroblasts	Heterozygous MID1 c.1801_1804delCTCC
			16/98 M-ctrl	M-ctrl	iPSCs	
			16/98 M-ctrl NPCs	M-ctrl	NPCs	
			16/98 M-ctrl Neurons	M-ctrl	Neurons	
1262/16	26 years	Female	1262/16	Fibro	Fibroblasts	Wild type
			1262/16 iPSCs	J-ctrl	iPSCs	
			1262/16 iPSCs	J2-ctrl	iPSCs	
			1262/16 NPCs	J-ctrl	NPCs	
			1262/16 NPCs	J2-ctrl	NPCs	
			1262/16 Neurons	J-ctrl	Neurons	
1179/17	~60 years	Female	1179/17	Fibro	Fibroblasts	Wild type
			1179/17 iPSCs	A-ctrl	iPSCs	

			1179/17 NPCs	A-ctrl	NPCs	
			1179/17 Neurons	A-ctrl	Neurons	

3.2 Methods

3.2.1 Cell culture

All cells were cultured at 37°C in a humidifier incubator supplied with 5% CO₂ unless otherwise specified. All cell culture procedures were conducted under a laminar working bench under sterile conditions.

3.2.1.1 Coatings

Coating of the plates is required for the proper adherence and to maintain the proper cell identity. Plates were coated for the culturing of iPSCs, NPCs, and neurons, as well as special coating were used during specific procedures.

3.2.1.2 Gelatine

Gelatine-coated plates were used for critical steps during fibroblast isolation from skin punch biopsy and during the reprogramming of iPSCs. Gelatine stock solution (2%) was pre-warmed and diluted 1:20 with pre-warmed PBS, and subsequently sterile-filtered. For the coating of 6-well plates, 1 ml of the solution was transferred into each well of the 6-well plate. The plates were incubated for 1 hour in a humidified incubator before use.

3.2.1.3 Matrigel

Matrigel is the name of the commercially available 3D basement membrane collected from the Engelbreth-Holm-Swarm (EHS) mouse sarcoma. Matrigel product is mimicking the laminin/collagen IV-rich extracellular matrix found in many tissues and it is widely used for culturing stem cells and primary cells. Prior use, Matrigel was thawed overnight at +4°C, aliquoted according to the dilution factor (LOT-specific), and stored at -20°C. Matrigel aliquot was diluted into 50 ml of chilled DMEM/F-12, mixed and incubated for 1 hour at +4°C. For the coating of 6-well plates, 2 ml of the solutions were aliquoted into each well of the 6-well plate and incubated for 1 hour at +37°C in humidified incubator prior use.

3.2.1.4 Geltrex

Geltrex is another commercially available 3D basement membrane obtained in similar way as for the Matrigel. Thanks to a different manufacturing process, the protein concentration is stable within batches. For the coating of 6-well plates, Geltrex was diluted 1:100 in 6 ml of pre-chilled DMEM/F-12, and 1 ml of the solution was dispensed in each well. Plates were used after 1 hour of incubation at +37°C in humidified incubator prior use.

3.2.1.5 Poly-Ornithine/Laminin

Poly-Ornithine/Laminin is a surface coating solution resulting from the combination of Poly-L-Ornithine and Laminin. Poly-Ornithine/Laminin coating is widely used for culturing different types of neuronal and neural stem cells cells, promoting neural cell adherence and differentiation. Poly-Ornithine stock solution (500 µg/ml) was diluted 1:50 in pre-chilled Borat buffer solution (150mM; pH 8.35). For the coating of 6-well plates, 1.5 ml were dispensed into each well. Plates were incubated over night at +37°C in a humidified incubator. On the second day of the coating protocol, Poly-Ornithine solution was removed and plates were washed three times with 2 ml of HBSS solution. Plates were then incubated for 3 hours at +4°C with a solution of 1 µg/ml Laminin in HBSS, and long-term stored (up to 6 months) at -20°C. Plates were thawed for at least 1 hour at +37°C in a humidified incubator prior use.

3.2.2 Cell culturing

3.2.3 Fibroblasts

Human Dermal fibroblasts from the M-, J- and A-line were used in the thesis. Fibroblasts were previously isolated from patient and healthy control (Thesis, Dr. Stephan Käseberg), while fibroblasts from an additional healthy individual (F-line) were isolated during the study.

3.2.3.1 Isolation

Dermal fibroblasts were isolated from a 4 mm skin punch biopsy by following a published protocol with small adjustments⁴⁵⁰. Skin biopsy was collected with medical procedure in the Human Genetic Department in Mainz, and delivered to the research lab in pre-chilled Fibroblast Extraction Media (FEM: DMEM +20% FBS +1% Pen/Strep). Skin biopsy was processed under laminar working bench by using a stereomicroscope. Biopsy was divided into 20-25 equally-sized pieces, and pieces were evenly distributed in a gelatine-coated six-well plate (see section 3.2.1.2), containing 800 µl of FEM and cultivate at +37°C with 5% CO₂ in a humidified incubator. 200 µl of FEM were added on alternate days to prevent the cells from drying out. After one week from biopsy processing, FEM was increased to 2 ml for each well, and replaced on alternate days. After 3-4 weeks, when fibroblasts migrated out the skin biopsy and reached full confluency, cells were detached and seeded on two T75 flasks by using the splitting procedure(see section 3.2.3.4). When cells reached complete confluency again, fibroblasts were transferred to three T175 flasks. Fibroblasts were stored in fibroblast freezing media (IMDM +15% FBS +1% Pen/Strep +10% DMSO) with the concentration of 1 million cells/ml by using the freezing procedure (see section 3.2.3.5). 1 ml of cell suspension was frozen for each cryovial. Cells were stored for 24-48 hours in insulated boxes at -80°C and then transferred to liquid nitrogen where they were long-term stored.

3.2.3.2 Thawing

Cells were thawed quickly by partially immersing the cryovial into +37°C water or by using hands. Once thawed, cells were transferred into a 15 ml conical tube containing pre-warmed fibroblast culture media, and subsequently centrifugated for 4 min at 200g. Supernatant was decanted and pellet was resuspended in fresh pre-warmed media. Solution was then transferred to a T75 flasks. Thawing procedure was further used for all different cell types by adapting the cell culture media and cell culture plate to the specific cell type.

3.2.3.3 Culturing

Human dermal fibroblasts were cultured in T75/T175 flasks containing fibroblast culturing media (IMDM +15% FBS +1% Pen/Strep). Cells were monitored every other day and cultivate at +37°C with 5% CO₂ in a humidified incubator. Cells were routinely checked for Mycoplasma contamination.

3.2.3.4 Splitting

When cells reached 80%-90% confluency, fibroblasts were detached and replated in a new flask in a splitting ratio between 1:2 and 1:10 ratio, depending on the specific need. For detaching the cells, cells were washed with PBS and incubated for 5 minutes with TrypLE enzyme at +37°C. Cells were collected with pre-warmed fibroblast culturing media and resuspended to ensure proper single-cell solution. Cells were counted by using the Neubauer chamber and desired number of cells were transferred into a new flask containing fibroblast culturing media.

3.2.3.5 Freezing

Cells were treated as described in the splitting section (see section 3.2.3.4). After assessing the number of cells, fibroblasts were centrifuged in a 15 ml tube for 4 minutes at 200g. Supernatant was removed and pellet was resuspended with fibroblast freezing media (IMDM +15% FBS +1% Pen/Strep +10% DMSO) with the final concentration of 1 million of cells/ml. 1 ml of the solution was aliquoted in cryovials and stored in insulated boxes at -80°C for 24-48 hours. Cryovials were then transferred into liquid nitrogen for long-term storage.

3.2.4 Induced Pluripotent Stem Cells (iPSCs)

iPSCs used in the thesis were generated by transducing human dermal fibroblasts with the commercially available CytoTune™-iPS 2.0 Sendai Reprogramming Kit by following the Feeder-Dependent approach. Reprogramming procedure was previously performed in the laboratory (Thesis, Dr. Stephan Käseberg).

3.2.4.1 Thawing

iPS cells were thawed as described in the fibroblast thawing procedure (see section 3.2.3.2) with small changes. After thawing, cells were transferred into 15 ml tube containing DMEM/F-12, while cell pellet was resuspended in iPSCs culturing media (mTeSR1 +1%Pen/Strep). iPS cells were then seeded into 6-well plates previously coated with Matrigel or Geltrex (see section 3.2.1.3, 3.2.1.4).

3.2.4.2 Culturing and cleaning of spontaneously differentiated cells

iPS cells were cultured in 6-well plates coated with Matrigel or Geltrex in iPSCs culturing media (mTeSR1 +1%Pen/Strep). Medium was changed daily with pre-warmed medium. During extensive culturing and after critical procedures such as reprogramming and electroporation, iPS cells can spontaneously differentiate and lose their pluripotency status. To ensure proper iPS starting material for the experiments, differentiated cells were manually removed from the plate. Manual removal of differentiated cells was performed under laminar working bench under sterile conditions. Differentiated cells were identified by using the stereomicroscope and removed by scraping the plate by using a flame-polished Pasteur pipette angled tilted and with a sharpened end. After removing the differentiated colonies, cells debris were removed by washing the well plate with DMEM/F-12 and replaced with culturing media.

3.2.4.3 Splitting

When iPS colonies were 70-80% confluent, cells were re-plated in a new Matrigel or Geltrex-coated 6-well plate. For detaching the iPS cells, a self-made enzyme-free solution called PBS-EDTA was used (500 ml PBS, 0.9 g NaCl, 250 µl 1M EDTA). Cells were washed twice with PBS-EDTA and incubate for 2 min with 1 ml of PBS-EDTA at room temperature. After the incubation time, PBS-EDTA was replaced with mTeSR1 and cells were scraped with cell scraper and cell clumps were collected by adding 2.5 ml of mTeSR1. Solution was resuspended until obtaining cell clumps of about 50-500 cells, and the desired amount of cell solution was seeded in a Matrigel or Geltrex-coated 6-well plate.

For the electroporation protocol, as well as for some differentiation protocols, iPS cells were needed as single-cell solution. For that reason, a different splitting procedure was applied. iPSCs were washed once with PBS and incubated with 1 ml of TrypLE enzyme or Accutase for 5 minutes at +37°C. Depending on the specific protocol, cells were collected with PBS or DMEM/F-12 and resuspended to obtain a single cell solution. After counting the cells using a Neubauer chamber, the desired number of cells was centrifuged for 4 minutes at 200g before proceeding with further procedures.

3.2.4.4 Colony picking

During fibroblast reprogramming, to ensure the clonal derivation of iPSCs, and to rescue heavily differentiated iPS lines, picking of iPS colonies was performed under laminar working bench under sterile conditions. iPS cells were treated as during the splitting procedure with PBS-EDTA. After incubation, 2 ml of mTeSR1 was added to the cells, and single iPS colonies were detached by simultaneously scraping and aspirating with a 200 µl pipette. The plastic tip was cut to provide a larger surface area for scraping the plate. iPS colonies were then transferred to a well of a Matrigel or Geltrex-coated 12-well plate.

3.2.4.5 Freezing

iPS cells were treated as described for the splitting procedure with PBS-EDTA. After incubation with PBS-EDTA, DMEM/F-12 was added to the cells and cells were detached by scraping the plate with a cell scraper. Cells were then centrifuged for 4 min at 200 g. Supernatant was decanted and pellet resuspended with iPSC freezing medium (mTeSR1 +20% KOSR +10% DMSO +1% Pen/Strep). Cell solution was transferred into cryovials and stored for 24-48 hours in insulated boxes at -80°C and then transferred to liquid nitrogen for long-term storage.

3.2.4.6 Genome editing

Genome editing of iPSCs were performed to introduce the same *MIDI* mutation (c.1801_1804delCTCC) found in the patient-derived M-line. CRISPR/Cas9 was employed to modify the J-line iPSCs before the start of my PhD thesis. A-line iPSCs were modified via CRISPR/Cas9 during my thesis with the help of Azza Soliman, Hala Ibrahim, and Verena Engelhardt. Guide RNAs (gRNAs) were previously designed by using the website tool (crispr.mit.edu) and cloned in the gRNA cloning vector⁴⁵¹ (gRNA_Cloning Vector was a gift from George Church (Addgene plasmid #41824)). Cas9 protein was included in the CAG-Cas9-Venus plasmid⁴⁵² (pU6-(BbsI)sgRNA_CAG-925 Cas9-venus- bpA was a gift from Ralf Kuehn (Addgene plasmid # 86986)). To insert the same patient mutation in wild type cell lines, a template containing the patient gene sequence was used to induce the homology- directed repair. A dsDNA break was induced by the use of the gRNA, and the template sequence was used to insert the patient *MIDI* sequence including the mutation. Cloning of the gRNAs and genome editing procedure of the iPSCs via electroporation were performed as described before (Thesis, Dr. Stephan Käseberg). Clones carrying the desired mutation were further expanded and used for further experiments.

3.2.5 Neural progenitor and stem cells (NPCs)

Neural progenitor and stem cells were employed in the study as the first cell line differentiated towards human neurodevelopment. NPCs were differentiated from several iPSC clones and lines, and successful differentiated cells were cultivated as stable NPC lines that maintains stem cell properties and being able to differentiate into neurons until passage 20.

3.2.5.1 Differentiation of iPSCs into NPCs

NPCs were obtained from the M-line iPSCs before I started my PhD thesis (Thesis, Dr. Stephan Kaseberg). M-line NPCs were differentiated by following an in-house protocol based on the generation of embryoid bodies and differentiation of neuronal rosettes from embryoid bodies attached to the plate. NPCs were obtained as proliferative cells by picking differentiated neuronal rosettes and dissociating them by incubating the neuronal rosettes with TrypLE enzyme and following the standard splitting protocol (Thesis, Dr. Stephan Kaseberg).

Due to inconsistency in the results obtained from the differentiation protocol described above, a second differentiation method was established based on a commercial kit using the PSC Neural Induction Medium (NIM). On day -1, iPSCs were detached by using the PBS/EDTA solution and replated on a Matrigel- or Geltrex-coated six well plate with different concentrations. On day 0, cells reaching the 15-25% of confluency were chosen for the NPC differentiation, and media was changed from mTeSR1 to Neural Induction Medium (NIM: Neurobasal, 2% Neural Induction Supplement, 1% P/S). Media was replaced with increasing volumes according to the differentiation protocol. On day 7 of neural induction cells reached full confluency and were replated following the manufacturer's protocol. Briefly, cells were washed with PBS and incubated with pre-warmed Accutase for 7 minutes at +37°C. Cells were collected with PBS, resuspended and filtered with a 100 µm strainer to avoid the presence of iPSC cell clumps. Cells were then centrifuged for 4 min at 300g, and pellet was washed a second time with PBS. After centrifugation, cells were resuspended with Neural Expansion Media (NEM: 49% Neurobasal, 49% Advanced DMEM, 2% Neural Induction Supplement, 1% P/S) supplemented with ROCK inhibitor (5µM) and seeded in a Geltrex-coated 6-well plate (500x10³ cells/well). After 24 hours, medium was exchanged with NEM without ROCK inhibitor and cells were monitored daily with medium changes every other day. Depending on the presence of falsely differentiated cells, cells were cultured with NEM in Geltrex-coated six-well plates (500x10³ cells/well), and replated when reaching confluency by using an adjusted protocol. Briefly, after washing the cells with PBS, cells were incubated with Accutase at room temperature for 3-4 minutes. After the incubation, Accutase was removed, and cells were gently washed with PBS to remove the detached cells containing the falsely differentiated cells. Cells were again incubated with Accutase for 4 min at +37°C and replating protocol was followed as described above. When reaching confluency, NPCs not containing falsely differentiated cells were replated using Accutase as described above and cultured on poly-Ornithine/Laminin-coated dishes with neuronal medium (NM: DMEM/F-12, 1% N2-supplement, 2% B27-supplement, 1% P/S) supplemented with FGF2 (20ng/ml).

3.2.5.2 Culturing

Successful differentiated NPCs were cultivated on poly-Ornithine/Laminin-coated dishes with neuronal medium (NM: DMEM/F-12, 1% N2-supplement, 2% B27-supplement, 1% P/S) supplemented with FGF2 (20ng/ml). NPCs used for the experiments were not replated more than 20 times to avoid aberration due to extensive culturing.

3.2.5.3 Splitting

NPCs were replated when reaching confluence, generally after 5-7 days after plating. Cells were washed with PBS and incubated with TrypLE™ Express for 5 min at +37°C. Detached cells were collected with NM media supplemented with 15-30% KOSR. Cells were resuspended thoroughly and centrifuged for 5 min at 200g. Cell pellet was resuspended in NM supplemented with FGF2 (20ng/ml) and cells were plated on poly-Ornithine/Laminin-coated 6-well plate (100-200x10³ cells/well).

3.2.5.4 Freezing

For long-term storage, NPCs were frozen and stored in liquid nitrogen. For freezing, cells were treated as described in the splitting procedure. Once the cells were resuspended with NM supplemented with 15-30% KOSR, cells were counted and centrifuged for 5 min at 200g. Cell pellet was resuspended with NM media supplemented with 10% KOSR and 10% DMSO (1x10⁶ cells/ml). 1 ml of cell solution was added per freezing tube, and stored at -80°C for 48 hours before long-term storage in liquid nitrogen.

3.2.5.5 Thawing

NPCs were thawed as described in the fibroblast thawing procedure with small changes. After thawing, cells were transferred into 15 ml tube containing NM media supplemented with 15-30% KOSR, while cell pellet was resuspended in NM supplemented with FGF2 (20ng/ml). NPC cells were then seeded into poly-Ornithine/Laminin-coated 6-well plate.

3.2.6 Neurons

NPCs were further differentiated into post-mitotic neurons by following an in-house protocol based on a commercial differentiation protocol. Neurons were used for further experiment after 35 days, when the cells acquired a neuronal differentiation status similar to post-mitotic neurons. The low cell density at the beginning of neuron differentiation allowed the initial growth of NPCs without reaching confluency of the plate. Due to the differentiation protocol, the presence of non-neuronal cells was detected, such as astrocyte cells.

3.2.6.1 Differentiation and culturing

NPCs were harvested as described in the splitting procedure. After centrifugation, cells were seeded in a low concentration onto poly-Ornithine/Laminin-coated 6-well plate ($30-90 \times 10^3$ cells/well) with NM media supplemented with FGF2 (20ng/ml). After 24 from the seeding of the cells, cells were washed twice with PBS and media was replaced with NM+VitA (DMEM/F-12, 1% N2-supplement, 2% B27+VitA supplement, 1% P/S). Cells were cultured in a humidified incubator at 37°C and 8% CO₂. Media was never replaced, but every 3-4 days fresh medium was added to the cells. When reaching a total volume of 10-12 ml each well, media was removed (20-50%) before adding fresh media. Neurons were harvested or further processed for experiments after 35 days of differentiation.

3.2.7 Cell culture methods

A series of protocols were established in the cell culture for a standard sample collection and sample preparation to ensure proper reproducibility using different cell lines and cell types.

3.2.7.1 Cell pellet collection

Cell pellets were collected for further genomic DNA or total RNA extraction from every cell type described above. Cells were harvested 16-18 hours after a full media change, harvesting the cells during their exponential growth phase. Neurons were harvested after 16-18 hours of replacing 50% of their media. To detach the cells from the plate, media was removed and cells were washed with PBS. PBS was replaced after washing step, and cells were manually detached with a plastic cell culture scraper. Detached cells were collected by adding additional PBS and cells were pelleted in a 15 ml tube for 4 min at 200g. Cell pellet was resuspended in fresh PBS and cell solution transferred to a 1.5 ml tube. Cells were centrifuged for 5 min at 4500 rpm, and cell pellet was dry frozen at -80°C until further use.

3.2.7.2 PFA fixation of adherent cells

Adherent cells were fixed with paraformaldehyde solution (PFA 4% in PBS) before further processing for immunostaining or RNA-FISH procedures. To obtain adherent cells for imaging experiments, cells were seeded on coverslips and growth on 12-well plates. Coverslips were coated in the same way as wells of the 6-well plate, if required for the specific cell type. Cells were fixed when reaching confluency, while neurons were seeded and differentiated for 35 days before fixation. When cells were ready to be processed, media was removed and cells were gently washed twice with PBS. After washing with PBS, PFA was added to the well of the 12-well plate and incubated 15-20 min for immunostaining, and 30 min for RNA-FISH. After incubation with PFA solution, cells were washed three times with PBS

and stored at +4°C with fresh PBS for further processing. For RNA-FISH, fresh PFA solution was prepared before cell fixation, and dehydration procedure was performed. After three washes with PBS, cells were dehydrated by incubation with increasing volumes of ethanol, 50%, 70%, and 100% for 1 minute. 100% ethanol was replaced with fresh 100% ethanol, and cells were stored at +4°C for up to 48 hours.

3.2.7.3 Cell counting

Cells were counted after splitting procedure for replating or for freezing. After single cell dissociation, 10 µl of cell solution was pipetted into a Neubauer counting chamber. Cells in all four corner squares were counted. Cell concentration was calculated based on the formula:

$$\text{Concentration of cells } \left(\frac{\text{cells}}{\text{ml}}\right) = \frac{\text{cell count}}{\text{number of squares counted}} \times 2,5 \times 10\,000$$

3.2.7.4 Karyotyping

Karyotype analysis was performed after iPSC reprogramming to exclude any chromosomal aneuploidies. When reached confluency, iPSCs were processed in the diagnostics' laboratory of the "Institute of Human Genetics" of the University Medical Centre Mainz by the technician Denise Seyler. Therefore, the protocol is not given in full details.

3.2.8 Molecular methods

3.2.8.1 Immunofluorescence staining

Immunofluorescence method was employed to detect the presence of expressed proteins in cells. Cells growth on a coverslip and previously fixed with PFA solutions were used for the staining. After 3 washes in PBS, cells were incubated with blocking solution (PBS, 5% BSA, 0.3% Triton) for 30 minutes at RT, followed by overnight incubation at +4°C with the primary antibody diluted in blocking solution. The following day cells were incubated three times with wash buffer (PBS, 0.3% Triton) for 10 minutes. Cells were then incubated for 1 hour in the dark with the secondary antibody diluted in blocking solution. From this moment on, cells were always handled in the dark if possible. Following the incubation, cells were washed three times with wash buffer for 10 minutes at room temperature, and briefly washed with dH₂O before transferred to glass slides. Coverslips were mounted by using 10 µl of mounting medium (Vectashield with the supplement of 0.5% DAPI). Images were acquired by using the Echo Revolve microscope and analysed in Fiji.

3.2.8.2 RNA-FISH procedure and imaging

To trace the allelic expression of X-chromosomal genes in a single cell resolution we employed the RNAScope technology (BioTechne) by following manufacturer's protocol with small adjustments. As described before, cells seeded on coverslips were fixed when reaching 80-90% of confluency with 1ml of freshly prepared PFA (4% in PBS) for 30 minutes at room temperature. After three washes with PBS, cells were dehydrated by incubation with increasing volumes of ethanol, 50%, 70%, and 100% for 1 minute. 100% ethanol was replaced with fresh 100% ethanol, and cells were stored at +4°C for up to 48 hours. Cells were rehydrated (70% and 50% ethanol for 1 minute) followed by incubation with PBS for 10 minutes. For the pre-treatment procedures, coverslips were covered for 10 minutes with RNAScope Hydrogen Peroxide solution and incubated for 10 minutes with RNAScope Protease III diluted 1:5 with PBS. Hs-XIST-C2 probes were mixed with Hs-MID1-O1-C3, Hs-GPM6B-intron-C3, Hs-NLGN4X-

O1-C3, or Hs-SPIN3-Intron-C3 according to manufacturer's protocol. For the co-detection of MID1

with NLGN4X and XACT, Hs-XACT-C1 were mixed with Hs-MID1-O1-C2 and Hs-NLGN4X-O1-C3 according to manufacturer's protocol. Coverslips were incubated with mixed probes for 2 hours at 40°C. Negative control samples were incubated with Probe Diluent. Additional negative control samples were treated with RNase A (100 ug/ul) for 1 hour at +37 °C prior probe incubation. The subsequent steps of amplification and detection signals were performed as described in the protocol. The Opal Dye 520 was used for Hs-XIST-C2 and Hs-XACT-C1 (1:750, excitation 494nm, emission 525nm), Opal Dye 570 for Hs-NLGN4X-O1-C3 (1:750, excitation 550nm, emission 570nm), and Opal Dye 690 for Hs-MID1-O1-C3, Hs-GPM6B-intron-C3, Hs-NLGN4X-O1-C3, Hs-SPIN3-Intron-C3 and Hs-MID1-O1-C2 (1:750, excitation 676nm, emission 694nm) as fluorophores. Samples were mounted with ProLong Glass Antifade Mountant. Fluorescence microscopy images were acquired using the VisiScope 5 Elements spinning disc confocal, built on a Ti-2E stand and equipped with a spinning disc unit (CSU-W1, 50µm pinhole, Yokogawa), and controlled by the VisiView software. Images were acquired with a 60x water immersion objective with an additional 2x objective, and a sCMOS camera and BC43 spinning disk confocal. 3D stacks of images (with a voxel size of x 0.0568, y 0.0568, 100 nm for VisiScope, and x 0.0609, y 0.0609, 100 nm for BC43) were acquired for each sample.

3.2.8.3 Quantification of RNA-FISH signals

At least 5 images for each experiment were acquired for the signal quantification. Images were analyzed using the Imaris Software (version 10). Manual quantification was conducted after setting the same threshold values for all pictures. *XIST*-positive nuclei were annotated as nuclei with a clear *XIST* cloud representing the inactive X-chromosome territory. Nuclei having at least one signal from the specific probe used were considered as positive nuclei for the specific probe. For the colocalization experiments between *XIST* signal and *MIDI*, *GPM6B*, *NLGN4X*, and *SPIN3*, nuclei having a single probe signal non co-localizing with the *XIST* signal were considered as monoallelic nuclei, while nuclei having two probe signals, with one of them colocalizing with the *XIST* signal, were considered as biallelic nuclei. Pre-mRNA signals that were overlapping, adjacent to, or in close proximity (distant 2.5 µm) of the *XIST* cloud were considered as co-localizing with the *XIST* cloud. For the co-localization experiments between *MIDI*, *XACT*, and *NLGN4X*, nuclei with a single *MIDI* signal colocalizing with *NLGN4X* and/or *XACT* were considered as monoallelic nuclei, while nuclei with two *MIDI* signals colocalizing with *NLGN4X* and/or *XACT*, or nuclei with two *MIDI* signals with one of them colocalizing with *NLGN4X* signal were considered as biallelic nuclei. Same criteria were applied for annotating the co-localization between pre-mRNA *MIDI* and *XACT/NLGN4X* signals.

3.2.8.4 DNA extraction

Genomic DNA was isolated from cell pellets frozen and conserved at -80°C. DNA extraction was performed by using self-made buffers and a protocol established before the start of my PhD thesis (Thesis. Dr. Stephan Kaseberg). The procedure was performed at RT. Cell pellet was resuspended in 300 µl of SE-buffer (0.439 g NaCl, 0,841 g Na-EDTA, and 100 ml of water). After resuspension, 30 µl of proteinase K (15 mg/ml) and 30 µl of 10% SDS solutions were added, and cells were lysate during overnight incubation at +37°C. The following day, 100 µl of 5M NaCl solution was added to the cell lysate, and samples were vortexed for 15 seconds. Samples were then centrifuged for 15 min at 6000g, and only the supernatant containing the genomic DNA was transferred into a new 1.5 ml tube, while pellet containing the cell debris was discarded. For DNA precipitation, 600 µl of 100% ethanol was added to the samples, and after vigorous shaking samples were centrifuged for 20 min at 14000g. Supernatant was removed and pellet was resuspended twice with 1 ml of 70% ethanol, following centrifugation for 5 min at 14000g. After second wash with 70% ethanol, pellet air dried to remove any

trace of ethanol, and resuspended with 20 μ l of dH₂O. Concentration of genomic DNA was determined using the Nanodrop™ One Spectral photometer. Genomic DNA was stored at -20°C for further analysis.

3.2.8.5 Total RNA isolation

Total RNA was isolated using the High Pure RNA Isolation Kit from Roche by following the manufacturer's instructions for Isolation of Total RNA from Cultured Cells; procedure was performed at RT. Cell pellet was resuspended with 200 μ l of PBS, and 400 μ l of Lysis/Binding buffer was added to each sample. After vortexing samples for 15 seconds, cell lysates were transferred to one of the supplied filter tubes, and filtered by centrifugation for 30 seconds at 8000g. The flowthrough was discarded and the filter containing the sample was treated for 15 min at RT with 10 μ l of DNase I diluted in 90 μ l of DNase incubation buffer. After the incubation, the filter was washed using 500 μ l of wash buffer 1 and centrifuged for 30 seconds at 8000g. Same washing of the filter was performed with wash buffer 2. Additional washing steps was performed with wash buffer 2, by pipetting 200 μ l of wash buffer 2 onto the filter and centrifuge at 13000g for 2 min. Only the filter was then transferred to a fresh 1.5 ml tube and RNA was eluted using 50 μ l of elution buffer and centrifugating the samples for 1 min at 8000g. The RNA concentration was assessed by using the Nanodrop™ One Spectral photometer. RNA was then stored at -80°C for further analysis.

3.2.8.6 Retro-transcription of total RNA and cDNA synthesis

To assess gene expression analysis, total RNA was retro-transcribed into cDNA by following two comparable protocols. Total RNA was transcribed into cDNA by using the PrimeScript™ RT Master Mix and the RevertAid First Strand cDNA Synthesis Kit. For the PrimeScript™ RT Master Mix, 500 ng of total RNA were used, and water was added to a final volume of 8 μ l. 2 μ l of the 5x master mix was added to the reaction and the final solution was incubated as shown in Table 16. After the cycler program, samples were diluted with dH₂O to a final RNA concentration of 5 ng/ μ l.

Table 16 Cycler program used for cDNA synthesis by using the PrimeScript™ RT Master Mix

Temperature	Time
37°C	15 min
85°C	5 second
4°C	forever

For the RevertAid First Strand cDNA Synthesis Kit, 500 ng of total RNA were used for a single reaction. 1 μ l of Oligo (dT) primers and 1 μ l of Random Hexamer primers, together with dH₂O, were added to the solution to a total volume of 12 μ l. Additionally, 4 μ l of 5x Reaction Buffer, 1 μ l of RiboLock RNase Inhibitor (20 U/ μ l), 2 μ l of 10mM dNTP Mix, and 1 μ l of RevertAid M-MuLV RT (200 U/ μ l), for a total volume of 20 μ l. Final solution was incubated as described in Table 17. After the cycler reaction, samples were diluted with dH₂O to a final RNA concentration of 5 ng/ μ l.

Table 17 cycler program used for cDNA synthesis by using the RevertAid First Strand cDNA Synthesis Kit

Temperature	Time
25°C	5 min
42°C	60 min

70°C	5 min
4°C	forever

3.2.8.7 Sendai Virus vector clearance

To investigate the viral vector presence in the iPSC samples, cDNA was used for a RT-PCR reaction. Two different PCR reactions were performed by using two pair of primers for each reaction (Table 8), specific for the genes contained in the viral vectors used for the reprogramming procedure (KOS, KLF4, and c-Myc), while SeV primers amplify a specific Sendai Virus genome sequence. SeV and KOS primer pairs were used together in one PCR reaction, while KLF4 and c-Myc primers in a second PCR reaction. PCR reaction was performed in a total volume of 25 µL. 10 µL of cDNA were mixed with 2,5 µL of 10x buffer, 1 µL of each primer pair (Table 8), 0,5 µL dNTPs, 0,2 µL FastStart™ Taq DNA Polymerase, 1 µL of GC-rich supplement and filled up with water to the total volume of 25 µL. Cycler amplification was performed by following the suggested program from StemCell and reported in table 18. Successful amplification of PCR products was performed by loading 12.5 µL of the samples on a 1.5% Agarose gel supplemented with EtBr.

Table 18 cycler program used for the SeV vector RT-PCR

Stage	Temperature	Time
Initial denaturation	94°C	2 min
Denaturation	94°C	30 seconds
Primer annealing	55°C	30 seconds
Elongation	72°C	30 seconds
Loop from denaturation to elongation 34 times		
Final elongation	72°C	5 min
Store of the samples	4°C	Forever

3.2.8.8 Allele-specific RT-PCR

Allele-specific RT-PCR was established for the amplification of the specific patient-derived *MIDI* mutation present in the M-line and introduced via CRISPR/Cas9 in the A- and J-line. The allele-specific RT-PCR consists of two separate reactions amplifying the wild type or the mutant *MIDI* allele, the latter containing a 4 bp deletion (c.1801_1804delCTCC). For each cDNA sample, two PCR reactions were performed by using the same forward primer annealing in the exon 8, and two reverse primers, one complementary to the wild type exon 9, the other to the mutant exon 9', as described by the cartoon in figure 10.

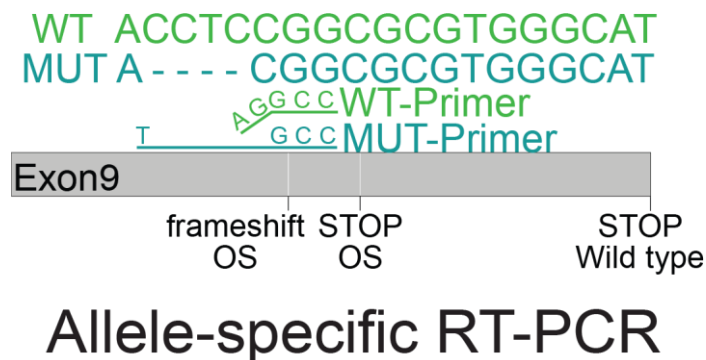


Figure 10. Allele-specific RT-PCR designed for the amplification of wild type and mutant *MIDI* transcripts. Cartoon representing the wild type (light green) and mutant (light blue) forward primers. The 4 bp deletion present in the mutant allele allows for the specific binding of the two primers and for the allele-specific amplification.

For the PCR reactions, 4 μ l of diluted cDNA were added to the PCR reaction containing 2,5 μ l of 10X buffer, 5 μ l of GC-rich buffer, 0,5 μ l of dNTPs, 10,8 μ l of dH₂O, 0,2 μ l of FastStart™ Taq DNA Polymerase, 1 μ l of the forward primer, and 1 μ l of the allele-specific reverse primer, either for the wild type or the mutant *MIDI* allele. Total volume of PCR reaction was 25 μ l, and PCR was performed as described in Table 19. PCR products were visualized by separation on a 1,5% agarose gel supplemented with EtBr. PCR products were loaded on the same gel with the same sample order to compare the expression between wild type and mutant *MIDI* allele.

Table 19 Cycler program for the allele-specific *MIDI* RT-PCR

Stage	Temperature	Time
Initial denaturation	94°C	2 min
Denaturation	94°C	30 seconds
Primer annealing	70°C	30 seconds
Elongation	72°C	40 seconds
Loop from denaturation to elongation 34 times		
Final elongation	72°C	5 min
Store of the samples	4°C	Forever

3.2.8.9 Quantification of Allele-Specific Expression by Pyrosequencing (QUASEP)

Quantification of Allele-Specific Expression by Pyrosequencing (QUASEP) assay was performed to investigate the allele-specific expression by quantify the relative expression of genomic variants. The generation of primers for the specific genomic variant was conducted by using the PyroMark Assay Design 2 software. The first step of the QUASEP assay consisted in the amplification of the region of interest where the genomic variant is located. Primer pair generated with the software were used to amplify the region of interest. As a preparation for the subsequent reaction, one of the primers was biotinylated. PCR was performed to amplify the region of interest, and genomic DNA and cDNA samples were measured as a technical triplicate to correct for technical errors. For a single PCR reaction, 10 μ l of cDNA or 1 μ l of genomic DNA were mixed with 5 μ l of 10x buffer, 1 μ l of dNTPs, 2,5 μ l of forward and 2,5 μ l of reverse primers, 28,6 μ l or 37,6 μ l of dH₂O, and 0,4 μ l of FastStart™ Taq DNA

Polymerase. PCR reaction was performed as described in Table 20. The same PCR reaction mix and program was used for the different QUASEP assay described in the thesis.

Table 20 Thermocycler program for QUASEP PCR reactions

Stage	Temperature	Time
Initial denaturation	94°C	2 min
Denaturation	94°C	15 seconds
Primer annealing	60°C	30 seconds
Elongation	72°C	45 seconds
Loop from denaturation to elongation 34 times		
Final elongation	72°C	5 min
Store of the samples	4°C	Forever

Successful amplification of the region of interest was confirmed by separating 7,5 µl of the PCR products on 1.5% agarose gel with EtBr. The presence of a single amplified fragment with its length corresponding to the amplified region was considered as positive result, and remaining samples were used for the Pyrosequencing reaction. Pyrosequencing was performed by following the manufacturer's protocol and using the Pyromark machine. Pyro Q CpG software was used to design the experiment and the loading scheme of the samples. The remaining 42,5 µl PCR products were mixed with 40 µl of self-made binding buffer and 4 µl of Streptavidin-Sepharose beads in a 96-well plate. To avoid the precipitation of the Streptavidin-Sepharose beads, 96-well plate was vortexed gently during the preparation of the remaining reagents. 40 µl of self-made annealing buffer was mixed with 2 µl of sequencing primer in a 96-well pyrosequencing plate. Vacuum Prep tool was used to process the samples before loading them into the pyrosequencer. The mixture containing PCR products, binding buffer, and Streptavidin-Sepharose beads was aspirated, leaving a dry pellet that was incubated for 5 seconds in 70% ethanol, denaturation buffer (0.2M NaOH), and wash buffer (10mM Tris-Acetate, pH 7.4). The dry pellet containing the PCR product was then resuspended into the solution prepared in the 96-well pyrosequencing plate, and incubated for 2 min at 80°C. Sequence cartridge was prepared by mixing the Pyromark Gold Q96 Reagents, the enzyme, the single nucleotide solutions, and the substrate mix for the pyrosequencing reaction. The pipetting scheme as well as the needed volumes were calculated by the Pyro Q CpG software. Pyrosequencing plate and loaded cartridge were placed into the sequencing machine, and sequencing performed. Results were analysed by using the Pyro Q CpG software and displayed as percentage values.

3.2.8.10 RT-qPCR

Quantitative reverse transcription polymerase chain reaction (RT-qPCR) was employed in the thesis to detect and quantify the presence of RNA in a sample while performing a PCR reaction. Solution was prepared by mixing 4 µl of cDNA together with 7.5 µl of TB Green® Premix Ex Taq™ II, 2.5 µl of dH₂O, and 1 µl of primer pair mix containing the forward and the reverse primers. Each PCR reaction was performed in a single well of a 96-well plate, and each cDNA sample was measured as technical triplicate to correct for technical errors. After preparing the PCR solutions, 96-well plate was loaded in the StepOnePlus System machine, and RT-qPCR reaction was performed as described in Table 21. Supplied software was used to design the experimental setup and the loading scheme of the samples, as well as for the initial analysis of the results.

Table 21 PCR program for RT-qPCR experiment

Stage	Temperature	Time
Initial denaturation	95°C	15 min
Denaturation	95°C	15 seconds
Primer annealing	60°C	30 seconds
Elongation	72°C	40 seconds
Loop from denaturation to elongation 34 times		
Final elongation	72°C	5 min
Calculation of Melt curve		

Data was analysed by using Microsoft Excel and R-Studio following the protocol based on the $\Delta\Delta C_t$ -method⁴⁵³. *GAPDH* was used as the reference gene.

3.2.9 RNA-sequencing

RNA-sequencing was performed to compare the transcriptome of iPSCs, NPCs, and neurons obtained from different female donors. M-ctrl, J-ctrl, and A-ctrl were used in the thesis for the RNA-sequencing experiments at different stages of the study. The sequencing of the M-ctrl and J-ctrl was performed in the diagnostics' laboratory of the "Institute of Human Genetics" of the University Medical Centre Mainz, while the sequencing of the A-ctrl line was performed in collaboration with the IMB Genomics facility. The RNA-sequencing analyses were performed with the help of Dr. Stephan Kaseberg, Dr. Stefan Diederich, Dewi Hartwich, and Dr. Hristo Todorov.

3.2.9.1 RNA-sequencing library preparation

For the M-ctrl and J-ctrl lines the library for RNA-seq experiments was prepared by using 5ng of total RNA and following the manufacturer's instructions of the Ovation® Solo RNA-Seq Library Preparation Kit. Dr- Stephan Kaseberg participated to the library preparation. The successful prepared library was then denatured prior sequencing by mixing 5µl of the 4nM library with 5µl of NaOH (0.2 M). After incubating for 5 min, 5µl of Tris buffer (200nM Tris-HCl, pH 7) were added to the solution. Finally, 985µl of prechilled HT1 were added before loading the solution into the sequencing machine. The sequencing was performed by using a NextSeq 500/550 as a paired-end run with 2 times 76 cycles and an expected output of 50 million reads per sample. The high coverage of reads per sample was necessary for allele-specific analysis.

For the A-ctrl line NGS library preparation was performed by using the Illumina's Stranded Total RNA Prep Ligation supplemented with Ribo-Zero Plus Kit by following the Stranded Total RNA Prep Ligation with Ribo-Zero Plus Reference Guide (Document # 1000000124514 v02 April 2021). 613 ng of total RNA was used for the library preparation for each sample, and 10 PCR cycles were used for subsequent amplification. To obtain purified library and exclude residual primer and adapter dimers, two purification steps were performed after PCR reactions. Quality of the libraries was analysed in a DNA 1000 chip on a 2100 Bioanalyzer and libraries were quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 4.0 Fluorometer. All samples were pooled in equimolar ratio and sequencing was performed by using NextSeq2000 P3 flow cell, PE for 2x 151 cycles plus 2x10 cycles for the dual index read and 1 dark cycle upfront R1 and Read 2.

3.2.9.2 RNA-sequencing data pre-processing

RNA-sequencing data was processed by Dr. Stefan Diederich and Dewi Hartwich from the diagnostics' laboratory of the "Institute of Human Genetics" of the University Medical Centre Mainz. For the M-ctrl and J-ctrl lines, sequencing reads were demultiplexed, and base call (BCL) files were converted into Fastq files using `bcl2fastq` conversion software (v2.17.1.14). Sequence adapters were trimmed and reads shorter than 6 base pairs were excluded from further analyses with `Cutadapt`⁴⁵⁴ (v0.18). Quality control on the trimmed reads was performed using `FastQC` (v0.11.7). The trimmed reads were then aligned to the human reference genome and transcriptome (hg19) using the `STAR aligner`⁴⁵⁵ (v2.5.3). PCR duplicates were removed from the mapped reads with the python script `nudup.py` (v2.3) from NuGen. For the A-ctrl line, adapter and quality trimming was carried out using the `BBDuk` tool from `BBMap` v38.86. Quality checks were done with `FastQC` v0.11.9. The reads were then mapped to the human reference genome (hg19) using `STAR` v2.7.10b, and a count table was generated with `featureCounts` v2.0.1 from the `Subread` package.

3.2.9.3 Allele-specific expression analysis (ASE) of the bulk RNA-sequencing data

Allele specific expression analysis was performed by Dr. Hristo Todorov with the supervision of Prof. Susanne Gerber. To perform allele-specific expression analysis (ASE), bulk RNA-seq data from the M-ctrl, J-ctrl and A-ctrl lines were processed together with the commercially available `NVIDIA Clara Parabricks Pipeline` (v3.5). The build-in RNA pipeline `rna_gatk` was used to align the fastq-files to the reference human genome (version hg19) with the `STAR aligner`⁴⁵⁵. After sorting the coordinate and marking duplicates of the resulting BAM file, a base recalibration step was performed before genome variant calling with `gatk Haplotype Caller`^{456,457}. ASE analyses were restricted to the X-chromosome by defining `-L chrX`. For joint genotyping, a genomic database was built with `gatk GenomicsDBImport` including all variants from all samples, followed by genotyping of the X-chromosomal variants that was performed with `gatk GenotypeGVCFs`. Gene names were annotated by using the `Ensembl Variant Effect Predictor`⁴⁵⁸ before transforming the resulting VCF file to a table format with `gatk VariantsToTable` for smaller file size and to facilitate further downstream analysis in R. R base functions were used to remove all the genomic variants that did not intersect known X-linked genes, multi- and monoallelic variants, variants with a quality score < 100, as well as intermediate and very long indels (>50 bp), as a quality control. Then the allelic ratio was calculated for each genomic variant site by dividing the number of reads mapping to the reference allele (termed No. Ref) by the total number of reads covering the variant site (termed No. Total). For further processing, a threshold of at least 5 reads in all samples was applied to consider a genomic variant (SNP or indel) to be expressed. The average allelic ratios were calculated for each cell type (iPSCs, NPCs, and neurons) in the sequenced M-ctrl, J-ctrl and A-ctrl lines, respectively, by summing up all the reads mapping to the reference allele across all replicate samples from the corresponding cell type and cell line, and dividing this number by the total number of reads covering the variant site. This analysis was performed only including variant sites that were covered by at least 20 reads on the group level. For the detection of reactivated genes, which are characterized by monoallelic expression in iPSCs but biallelic expression in NPCs and/or neurons, first all variant sites with an allelic ratio <0.025 or >0.975 in iPSCs were identified. The allelic ratio was then converted to an estimate of the probability of expression \hat{p}_i from the inactive chromosome X_i using the following formula:

$$\hat{p}_i = \begin{cases} \frac{No. Ref}{No. total} & \text{if } No. Ref < No. Alt \\ 1 - \frac{No. Ref}{No. total} & \text{otherwise} \end{cases}$$

A one-sided binomial test was employed to investigate if p_{xi} is significantly greater than 0.025, indicating statistically significant expression from the inactive X-chromosome. P-values were adjusted for multiple comparisons by using the Benjamini-Hochberg method implemented in the stats R package (v4.0.2). Results with an adjusted P-value < 0.01 were considered statistically significant, indicative of the presence of a reactivated variant site. Genomic variant sites were considered as full-escapee of X-chromosome inactivation if p_{xi} is significantly greater than 0.1 in all cell types (iPSCs, NPCs, and neurons). For the detection of late-silenced ASE sites, all variants with $p_{xi} < 0.025$ in NPCs and neurons were identified, and then used one-sided binomial tests to examine if Xi expression was significantly greater than 0.025 in iPSCs.

3.2.9.4 Estimation of biallelic expression of X-linked genes

Based on the allele-specific expression analysis on the X-linked genomic variants, allele-specific gene expression was calculated. For genes that intersected only one ASE genomic variant site, the Xi expression for the respective variant was used as an overall estimate for the gene biallelic expression. If a gene contained more than one type of biallelically expressed genomic variants, the gene was manually assigned to a single category (reactivated, escapee or late-silenced) by following a predefined set of rules: 1) the gene was assigned to the category with the highest number of biallelic genomic variants 2) if a gene covered the same number of biallelic genomic variants from different categories, then SNPs were considered as more reliable than indels, and genomic variants of the same type with a higher coverage were preferred. After assigning the X-linked genes to a unique category, we used the metaprop function from the meta-R package (v4.18.1) to calculate gene-level estimates of biallelic expression from values for individual genomic variants by using an inverse-variance method with logit-transformed proportions of Xi expression. 99% confidence intervals of the biallelic expression, based on a normal approximation, were also obtained with the same metaprop function. To further improve the biallelic expression estimation, manual curation of the biallelically expressed variant sites was performed. All predicted biallelic variant sites were subjected to manual inspection, founding several false positive variants. 22 variants from M-ctrl, 14 variants from J-ctrl and 59 variant sites from the A-ctrl line were flagged as false positives and excluded from the analysis. In most cases, the predicted biallelic expression for these sites was due to all but one replicate expressing the reference allele and only one the alternative (or vice versa). These positions were then excluded from the analysis because not matching with any biological explanation for the strange pattern of biallelic expression.

3.2.9.5 Generation of X-chromosome ideograms

The X-chromosome ideogram showing the cytogenetic location of all the reactivated, full-escapee and late-silenced genes was produced by using the karyoploteR package (v1.4.1) based on the hg19 genome assembly. Gene coordinates were obtained from the Ensembl annotation.

3.2.9.6 Overlap analysis with the known escape genes in humans and mouse

The overlap of reactivated, full escape and late-silenced genes identified in this study with known

constitutive and facultative escape genes in humans was performed using escape status assignment performed in two previous studies^{301,459}. To calculate an expected value, 1000 random sampling

iterations were performed by annotating the number of genes in each category (reactivated: $n=95$; full escapees: $n=82$; late-silenced: $n=29$) from all X-linked genes annotated in Ensembl, and determining the overlap with the previously identified escape genes. The enrichment was calculated by dividing the observed overlap by the mean value for the expected overlap from the 1000 random sampling iterations. The significance of the observed overlap was obtained using the cumulative distribution function of the normal distribution. Similar analysis was performed to study the overlap of the identified escape genes with known escape genes in mouse. To assess the escape status of orthologue genes of the identified reactivated, full escape and late-silenced genes in the mouse, a recently published study by Hauth et al., 2024 was employed which analysed the literature so far and described a consensus allele-specific expression status of X-linked genes, dividing escape genes in constitutive, variable and NPC-specific escape genes.

3.2.9.7 Enrichment analysis of Neurodevelopmental Disorder-associated genes

The overlap of reactivated, full escape and late-silenced genes with NDD genes was studied to understand their role in neurodevelopment and NDD phenotypes. The recently published GeneTrek database was used, containing high-confidence NDD-associated genes⁴⁴². Then, the expected overlap of biallelically expressed genes with NDD genes was calculated by randomly sampling the 513 X-chromosomal genes reported in the GeneTrek database, also expressed in the NPCs and neurons in our study (>10 reads in each sample following DESeq2 normalization of the data). In every of the 1000 random iterations, the number of genes present in the respective category was randomly picked (reactivated: $n=95$, late-silenced: $n=29$, escapee: $n=82$), and the number of NDD associated genes was determined in each iteration. The mean value over all iterations was determined and considered as the expected number of NDD-associated genes. The enrichment was then calculated as the ratio between the observed and expected number of NDD-associated genes in each class. The significance of the observed overlap in each class was calculated using the cumulative distribution function of the normal distribution.

3.2.9.8 Cluster analysis of biallelically expressed genes

The distribution along the X-chromosome of the identified escape genes was investigated by looking at the distribution of the different types of biallelically expressed genes on the X-chromosome. The mean pairwise distance between transcription start sites for reactivated, escapee, and late-silenced genes, was calculated respectively. We tested the probability of observing a lower mean pairwise distance in our analysis than observed by chance by comparing the actual distance to a background random distribution of pairwise distances for 1000 iterations of randomly selected X-linked genes which were expressed in NPCs and neurons in our bulk RNA-seq data (>10 reads in each sample following DESeq2 normalization of the data). The P -value calculated for the observation of a lower mean pairwise distance by chance was obtained using the cumulative distribution function of the normal distribution.

3.2.9.9 Protein-protein interaction network analysis of the reactivated genes

The construction of a protein-protein interaction (PPI) network of the reactivated genes was performed by using the data provided by the human binary reference interactome (HuRi)⁴⁶⁰ available at <http://www.interactome-atlas.org/>. 54 protein-coding reactivated genes were included in the HuRi portal. Those proteins were used to create the network. To construct a neural differentiation-specific PPI, only genes that were expressed in all NPC and neuron samples in our bulk RNAseq data (>10 reads in each sample following DESeq2 normalization) were used for the interaction network. The igraph package (v1.2.6) in R was used to visualize the network. GO term analysis of the PPI network of the reactivated genes and NPC- and neuron-specific genes was performed with the clusterProfiler package

(v3.16.1) using an adjusted *P*-value cut-off of 0.05. Redundant GO terms were combined with the simplify function.

3.2.9.10 Chromatin state sex-specific analysis of the reactivated and silenced genes

To investigate the relationship between X-chromosome reactivation and chromatin epigenetic state, data from the 15-state chromatin model from the Roadmap Epigenomics Consortium was used⁴⁶¹. In particular, the female (sample E082) and male (sample E081) fetal brain samples were employed for the analysis. The annotation files for the epigenomes were downloaded from https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html, and the ChromDiff tool⁴⁶² was used to estimate the percentage of each gene corresponding to each of the 15 chromatin states in the two epigenome samples. The chromatin state profiles of the 83 reactivated genes, included in the GENCODE annotation files distributed with the ChromDiff tool, were compared against the profiles of X-chromosomal genes expressed in NPCs and neurons (>10 normalized reads in all samples) but not detected as biallelically expressed in the thesis and not being previously reported as escapee genes^{301,459}. One-sided Wilcoxon rank sum tests were performed to investigate the enrichment of the epigenomic states in the reactivated genes as well as genes not detected as biallelically expressed. This analysis was performed separately in the female and male fetal brain samples. *P*-values were adjusted for multiple comparisons with the Benjamini-Hochberg method and an epigenomic state was considered enriched in the respective gene category if the adjusted *P*-value was <0.05.

3.2.9.11 Sex-biased tissue-specific expression of reactivated, late-silenced and full-escapee genes

The tissue-specific sex-bias expression of the reactivated, full-escapee, and late-silenced genes was investigated *in vivo* by extracting the information from the study by Oliva and colleagues⁴⁵⁹ which investigated differences of gene expression levels between males and females across 44 human tissues from GTEx V8. The distributions of the number of tissues with a female bias was calculated for each of our gene categories compared to inactive genes using Fisher's exact test.

3.2.9.12 Disease-ontology enrichment analysis

Disease ontology enrichment analysis was performed for the total of the reactivated genes identified using the DOSE R package v3.28.2 with an adjusted *p*-value cut-off of 0.05. X-linked genes from the Ensembl annotation were used as the reference background for this analysis.

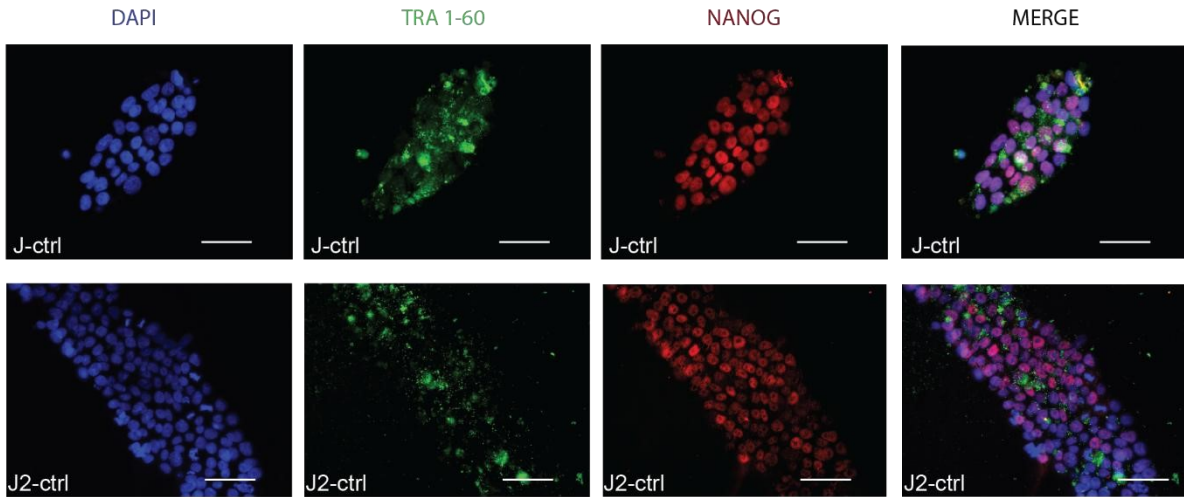
4 Results

4.1 Characterization of cell lines used for the study

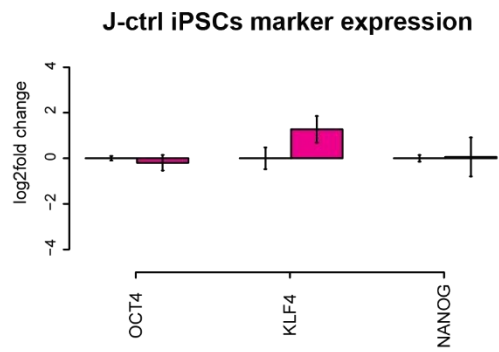
4.1.1 Induced pluripotent stem cells: confirmation of pluripotency identity and clonal status

Three female fibroblast lines were used in the study to investigate the expression status of escape genes during neural differentiation (Table 15). Human dermal fibroblasts were previously isolated from skin biopsy and reprogrammed into induced pluripotent stem cells (iPSCs) using an established workflow (see section 4.2.3.1). Confirmation of the fibroblast cell type identity and successful reprogramming into iPSCs were previously reported in a former lab member's PhD thesis (Thesis, Dr. Stephan Käseberg) for two female fibroblast lines, 16/98 and 1179/17, reprogrammed respectively into M-line and A-line iPSCs. The 1262/16 fibroblasts were isolated and reprogrammed into J-line iPSCs by following the same approach. Successful reprogramming of the 1262/16 J-line iPSCs was confirmed for two iPSC clones, J-ctrl and J2-ctrl, isolated during the reprogramming process and used for further experiments during the PhD thesis. The characterization followed the established workflow for previous iPSC lines characterization (Figure 11, Thesis, Dr Stephan Käseberg) by using RT-qPCR and immunostaining. Immunostaining confirmed the expression of the NANOG transcription factor in iPSC nuclei and TRA1-60 in the cytoplasm of the pluripotent cells, verifying the protein expression of pluripotency markers in J-ctrl and J2-ctrl iPSCs (Figure 11.A). Expression of pluripotency markers *NANOG*, *OCT4*, and *KLF4* was compared between the previously characterized A-ctrl iPSCs and the J-ctrl iPSC clone (Figure 11.B). Bar plot results showed the gene expression levels of the iPSC marker genes in J-ctrl iPSCs compared to A-ctrl iPSCs, indicating similar gene expression levels between A-ctrl and J-ctrl and confirming the high expression levels of pluripotency genes in J-ctrl. Taken together, these data confirmed the successful reprogramming of 1262/16 fibroblast cells into J-ctrl iPSCs (Figure 11.A, B). Karyotype analyses were performed to exclude the presence of aberrant chromosome rearrangements that can arise during reprogramming (Figure 11.C). The karyotype analysis was done in collaboration with the Diagnostic Human Genetics laboratory at the Unimedical Centre Mainz and was performed and analysed by Denise Seiler. The karyotypes of the J-ctrl clones used in the study showed no abnormalities in the number and composition of chromosomes. Female sex was confirmed for the two clones, possessing two X-chromosomes. The iPSC lines were reprogrammed in our laboratory by using the CytoTune™-iPS 2.0 Reprogramming System, based on a modified form of Sendai Virus (SeV) for the delivery of the key reprogramming transcription factors, *KLF4*, *c-MYC*, *OCT3/4*, and *SOX2*, sufficient for the proper resetting of terminally-differentiated somatic cells. Sendai Virus (SeV) vectors used in this protocol leads to transient expression of the transcription factor genes. Due to exclusive cytoplasmic replication and the absence of reverse transcription, integration of recombinant nucleic acid sequences into the cell's genome is not expected. Before using the J-ctrl and J2-ctrl iPS clones for further experiments, the complete clearance of the viral vector was confirmed to exclude any possible influence of the SeV vectors on the results. Generally, iPS clones were SeV vector-free after the first 10 passages. To further confirm the complete removal of the modified Sendai Virus, the temperature-sensitive mutations were employed by cultivating the cells at +39°C, and RT-PCR was performed (see section 4.2.8.7, Figure 11.D). Results showed the successful clearance of the SeV vectors in the J-ctrl and J2-ctrl iPSC clones tested. The +37°C, +39°C iPS sample, as well as the positive and negative control are divided into two gels, where the first consists on the detection of SeV and KOS vectors, while the second on *KLF4* and *c-Myc*. Positive controls containing the SeV vectors was used as comparison with the iPSC samples. The comparison between the positive control and the iPSC sample cultured at +39°C clearly showed the absence of the SeV vectors, confirming the complete clearance of the SeV vectors from the iPSC J-ctrl and J2-ctrl clones (Figure 11.D).

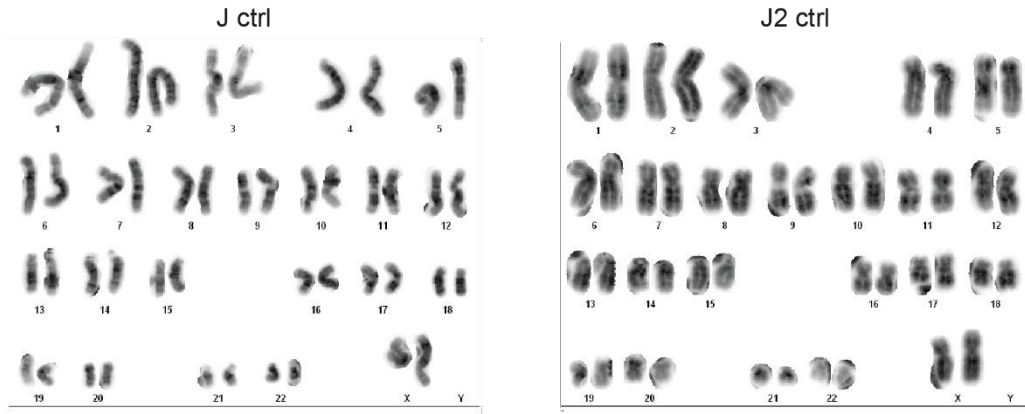
A



B



C



D

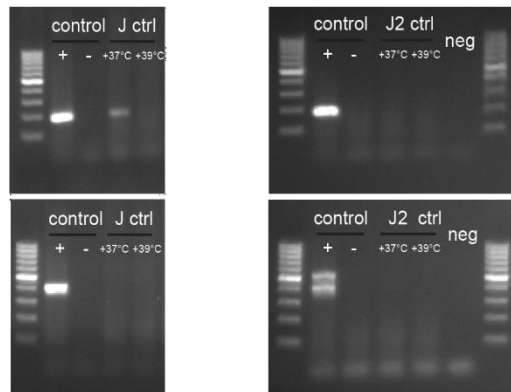


Figure 11. Identity confirmation of J-ctrl iPSCs. Successful reprogramming of J-ctrl and J2-ctrl iPSC clones was characterized by immunostaining and RT-qPCR analysis. A) Representative pictures of immunostaining for NANOG and TRA 1-60 pluripotency proteins. Scalebar = 50 μ m. B) Bar plot showing the relative expression of pluripotent marker genes *OCT4*, *KLF4*, and *NANOG* in J-ctrl replicates compared to A-ctrl measured via RT-qPCR. n=4 for A-ctrl and J-ctrl replicates. C) Karyotype analysis of female J-ctrl iPS cells. Clones used in the study, J-ctrl and J2-ctrl showed no chromosomal aberration and two X-chromosomes. D) Agarose gel showing the results of the RT-PCR for the Sendai Virus vectors. For each sample, two RT-PCR reactions were performed with two sets of primers for each reaction. Samples were loaded in the gel following the order: upper gel the SeV (181 bp) + KOS (528 bp) RT-PCR products, and lower the KLF4 (410 bp) + c-Myc (532 bp) RT-PCR products.

To investigate the expression of X-linked genes from the inactive X-chromosome, clonal iPSCs with skewed X-chromosome inactivation were used in this study. Quantification of Allele-Specific Expression via Pyrosequencing (QUASEP) assay was employed to confirm the clonal nature of the iPSCs. The three control lines, M-ctrl, J-ctrl, and A-ctrl, share a heterozygous SNP present in the 3' UTR of the X-linked gene *ZNF185* (rs3827418, chrX:152971838 C > G). The *ZNF185* gene is located in the Xq28 region and is reported to be subject to X-chromosome inactivation. The monoallelic or biallelic expression of the heterozygous SNP was tested in the M-ctrl, J-ctrl and A-ctrl lines (Figure 12.A, B, C). In genomic DNA (gDNA), QUASEP analysis confirmed the presence of the heterozygous SNP in the M, A and J cell lines. The relative abundance of the C nucleotide was 56% and 43% for G in M-ctrl genomic DNA, 43% and 57% in J-ctrl genomic DNA, while in A-ctrl DNA, the relative abundance of the C nucleotide was 37.5% and 62.5% for G. Fibroblasts showed biallelic expression of the heterozygous SNP, with the relative abundance of C at 73.32% and G at 26.68% for the M-line, 13.46% for C and 86.54% for G in the J-line, and 47.25% for C and 59.5% for G in the A-line. iPSCs showed a strictly monoallelic expression of either the C nucleotide (95.025% in the M-ctrl) or the G nucleotide (99.5% in the J-ctrl and 99.67% in the A-ctrl), depending on the iPS clone, thus confirming the clonal nature of the iPS clones. Additionally, iPS M-clones carried a heterozygous 4 base pair deletion in the *MIDI* gene, and allele-specific RT-PCR for *MIDI* wild type and mutated transcripts was performed. Agarose gel analysis of the PCR products confirmed a monoallelic expression of the *MIDI* gene in iPS M-clones (Figure 12.D). The presence of the wild type *MIDI* band in male control fibroblasts confirmed the specific primer amplification of either the wild type or mutant *MIDI* transcript. Female 16/98 M-line fibroblasts showed biallelic expression of the wild type and mutant *MIDI* alleles, consistent with the mixed population of fibroblast cells with random X-chromosome inactivation present in in vivo tissues. iPS M-clones showed a strict monoallelic *MIDI* expression of either the wild type (M-ctrl) or the mutant (M-OS het) *MIDI* isoforms. These results further confirmed the skewed X-chromosome inactivation status of the iPS M-clones.

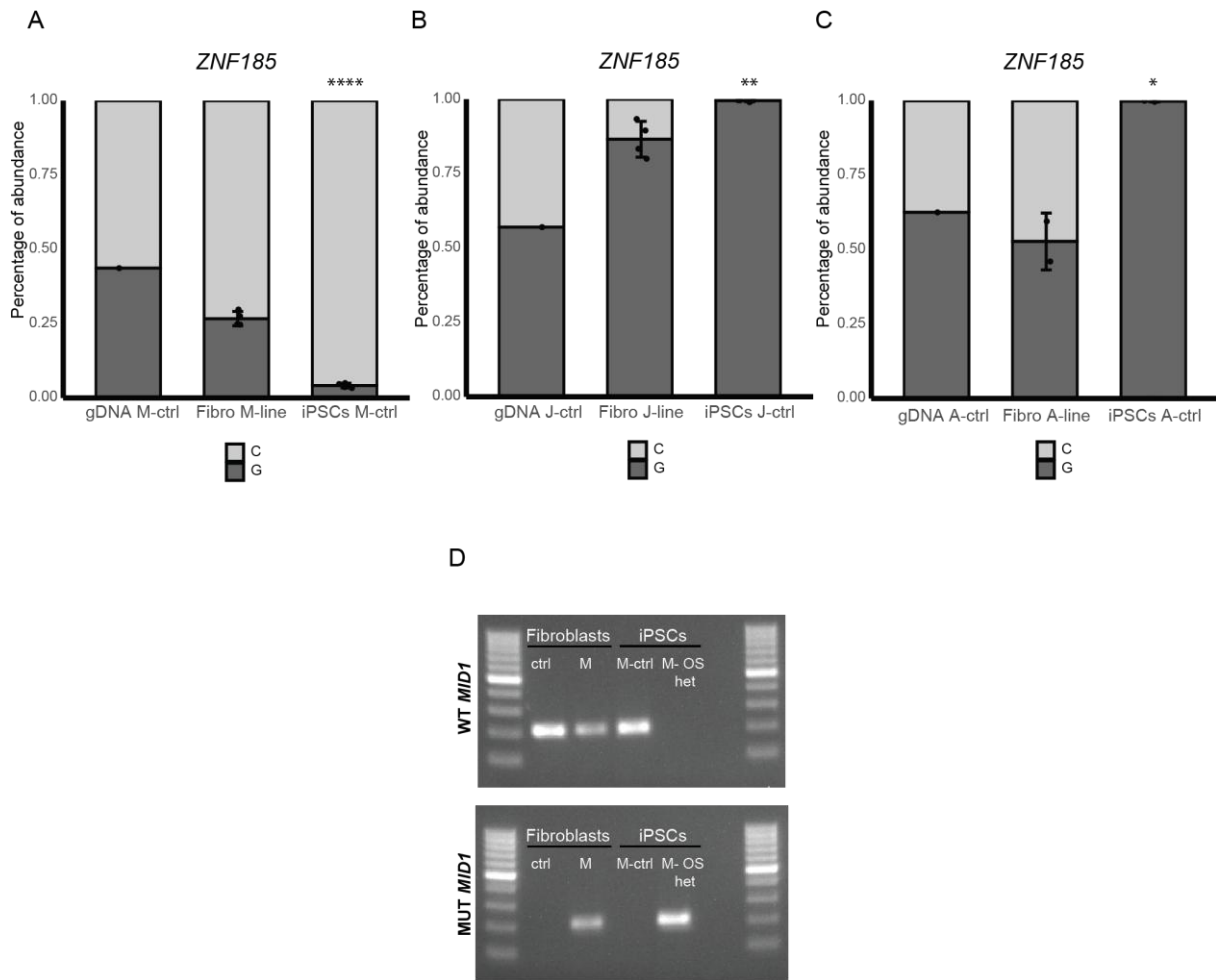


Figure 12. Confirmation of the clonal status in iPSCs. QUASEP assays (A, B, C) were used to assess the monoallelic or biallelic expression of a heterozygous SNP found in the X-linked *ZNF185* gene. A-B) n=1 for gDNA, n=4 for fibroblasts and iPSCs. C) n=1 for gDNA, n=2 for fibroblasts and iPSCs. Fibroblasts and iPSCs were compared by using One-way ANOVA followed by Tukey's multiple comparison test. Significance is assigned based on the adjusted P values: P<0.05 *, P<0.01 **, P<0.0001 ****. D) Allele-specific RT-PCR for MID1 transcripts. Ctrl fibroblasts were derived from healthy male donor and M-line fibroblasts were used as control.

4.1.2 Neuronal differentiation of iPSCs

4.1.2.1 Neural progenitor cells (NPCs) differentiation

M-ctrl, J-ctrl, and A-ctrl iPSC clones were differentiated into neural progenitor cells (NPCs) following the manufacturer's instruction of a published commercial kit (see section 4.2.5.1). Successful differentiation was confirmed by morphological analyses, RT-qPCR and immunostaining analyses. Confirmation of M-ctrl neural progenitor cell identity was previously reported (Thesis, Dr. Stephan Käseberg). RT-qPCR results showed a downregulation of iPSCs markers and a concomitant upregulation of NPCs markers in J-ctrl and A-ctrl NPC samples compared to iPSC controls (Figure 13). *OCT4*, *KLF4*, and *NANOG* pluripotency markers were downregulated in NPCs compared to iPSCs, while *SOX2*, *NESTIN*, and *PAX6* neural stem cell markers were upregulated in NPCs. The expression of the same NPC markers was confirmed at protein level via Immunostaining (Figure 14).

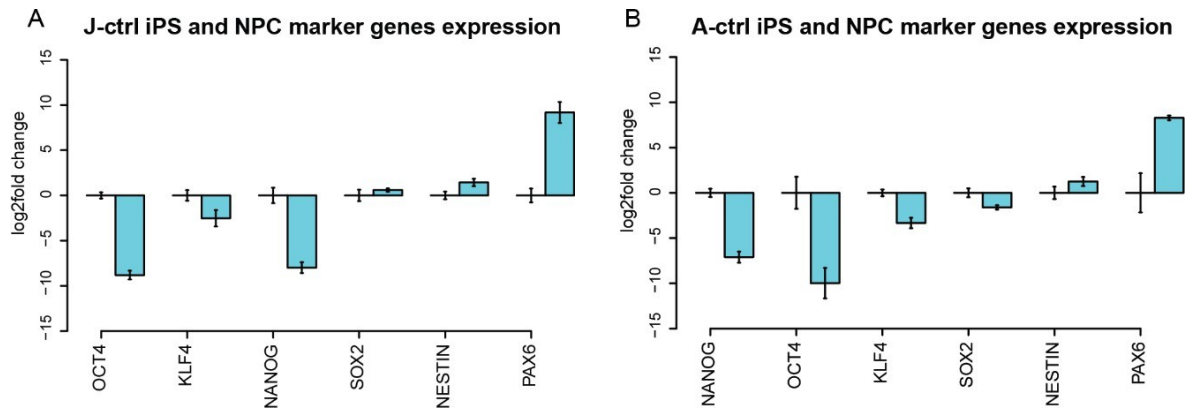


Figure 13. Confirmation of NPC cell identity via RT-qPCR. Pluripotency (*OCT4*, *KLF4*, and *NANOG*) and neural stem cell (*SOX2*, *NESTIN*, *PAX6*) marker gene expression was investigated in iPSCs and NPCs differentiated from them via RT-qPCR. J-ctrl (A) and A-ctrl (B) showed similar expression patterns of investigated genes. n=4 for J-ctrl; n=3 for A-ctrl.

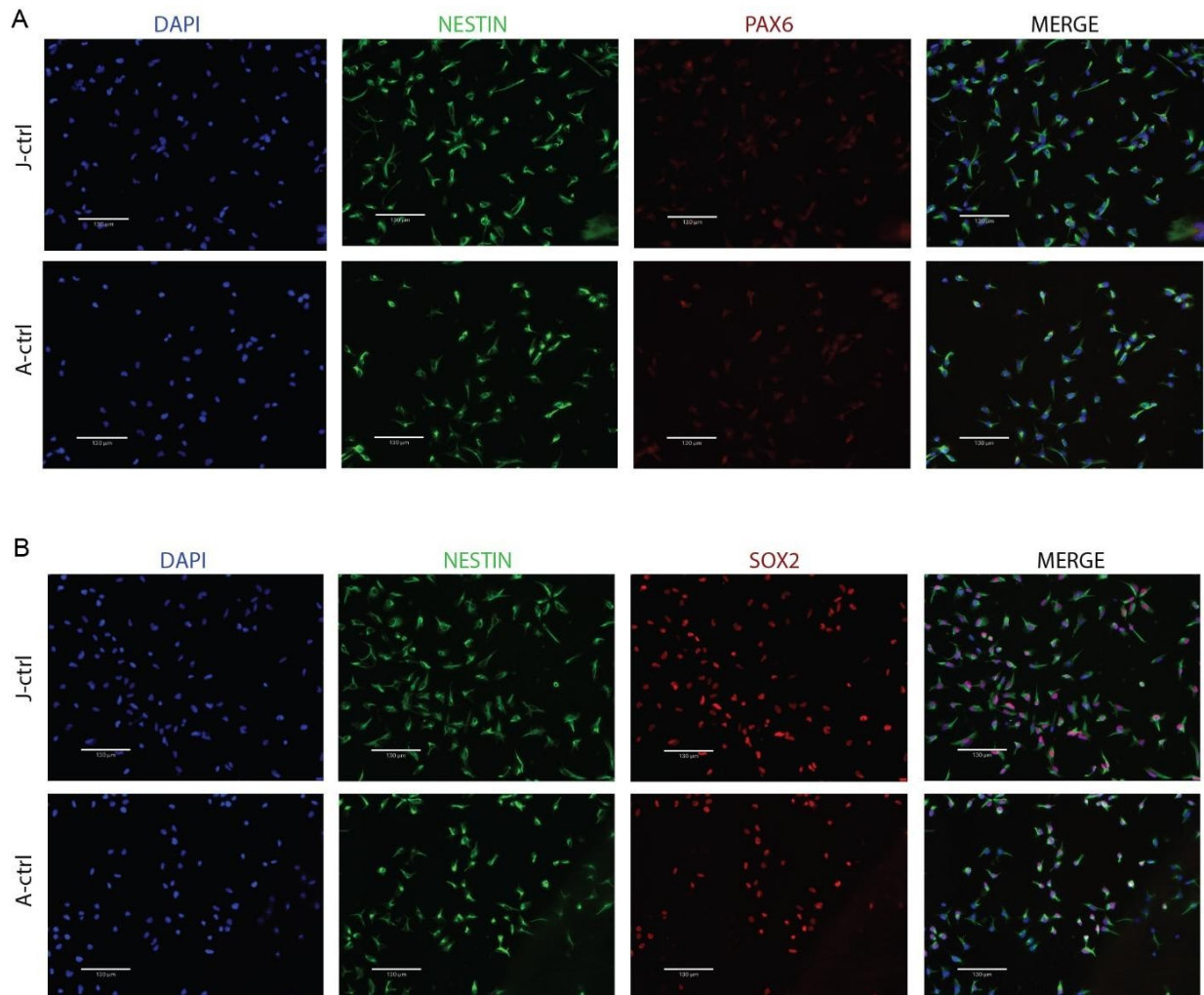


Figure 14. Confirmation of NPC cell identity via immunostaining. Protein expression of neural stem cell markers was confirmed via Immunostaining. A) J-ctrl and A-ctrl NPCs were positive for Nestin and PAX6 proteins. B) Expression of SOX2 was confirmed in Nestin-positive cells. Scale bar: 130 μ m.

4.1.2.2 Neuronal differentiation of NPCs

NPCs were further differentiated into neurons following an in-house established protocol (see section 4.2.6.1 and Thesis, Dr. Stephan Käseberg). M-ctrl neurons were previously characterized (Thesis, Dr. Stephan Käseberg). Characterization of J-ctrl and A-ctrl neurons was performed in this study. NPCs were seeded at low density and differentiated into neurons over 35 days with the addition of retinoic acid to the media. Successful differentiation was confirmed by immunostaining and RT-qPCR, which showed an upregulation of neuronal markers. RT-qPCR analysis indicated an upregulation of neuron-specific *TUBB3*, *MAP2*, and *TAU* markers in neurons compared to NPCs, along with a modest downregulation of NPC-specific *SOX2* and *NESTIN* markers (Figure 15). Notably, *PAX6* showed an upregulation of expression in both cell lines, suggesting the concomitant presence of non-fully differentiated cells. Immunostaining for the neuronal marker TUBB3, in combination with either MAP2 or TAU, showed co-localization of the proteins in axonal-structures (Figure 16). This confirmed the protein expression of neuronal markers in neuronal-like cells.

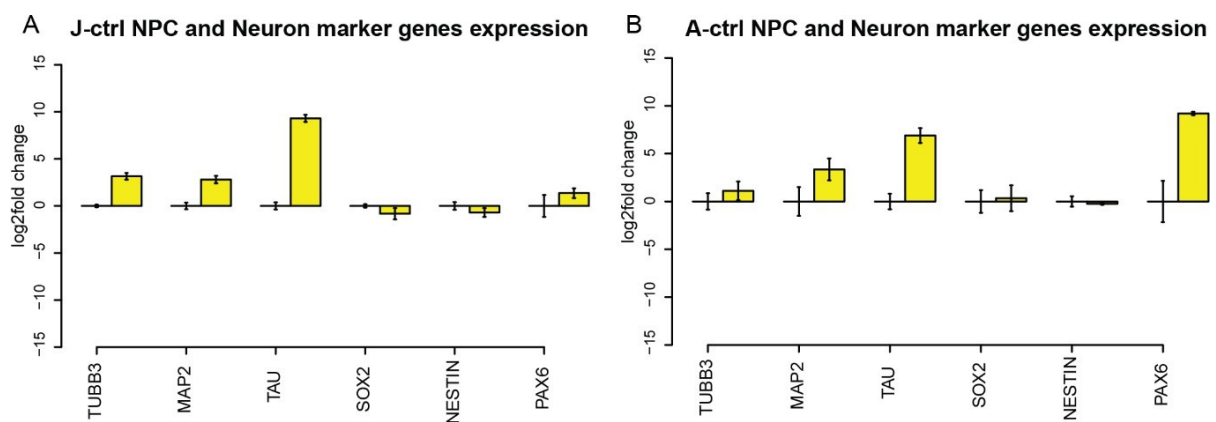


Figure 15. RT-qPCR results confirmed neuronal cell identity. Expression of NPCs (*SOX2*, *NESTIN*, and *PAX6*) and neuronal (*TUBB3*, *MAP2*, and *TAU*) marker genes in J-ctrl (A) and A-ctrl (B) showed similar expression pattern, with upregulation of *TUBB3*, *MAP2* and *TAU* neuronal markers and the *PAX6* NPC marker, and similar expression or modest downregulation of *SOX2* and *NESTIN* NPC markers. n=4 for J-ctrl. n=3 for A-ctrl.

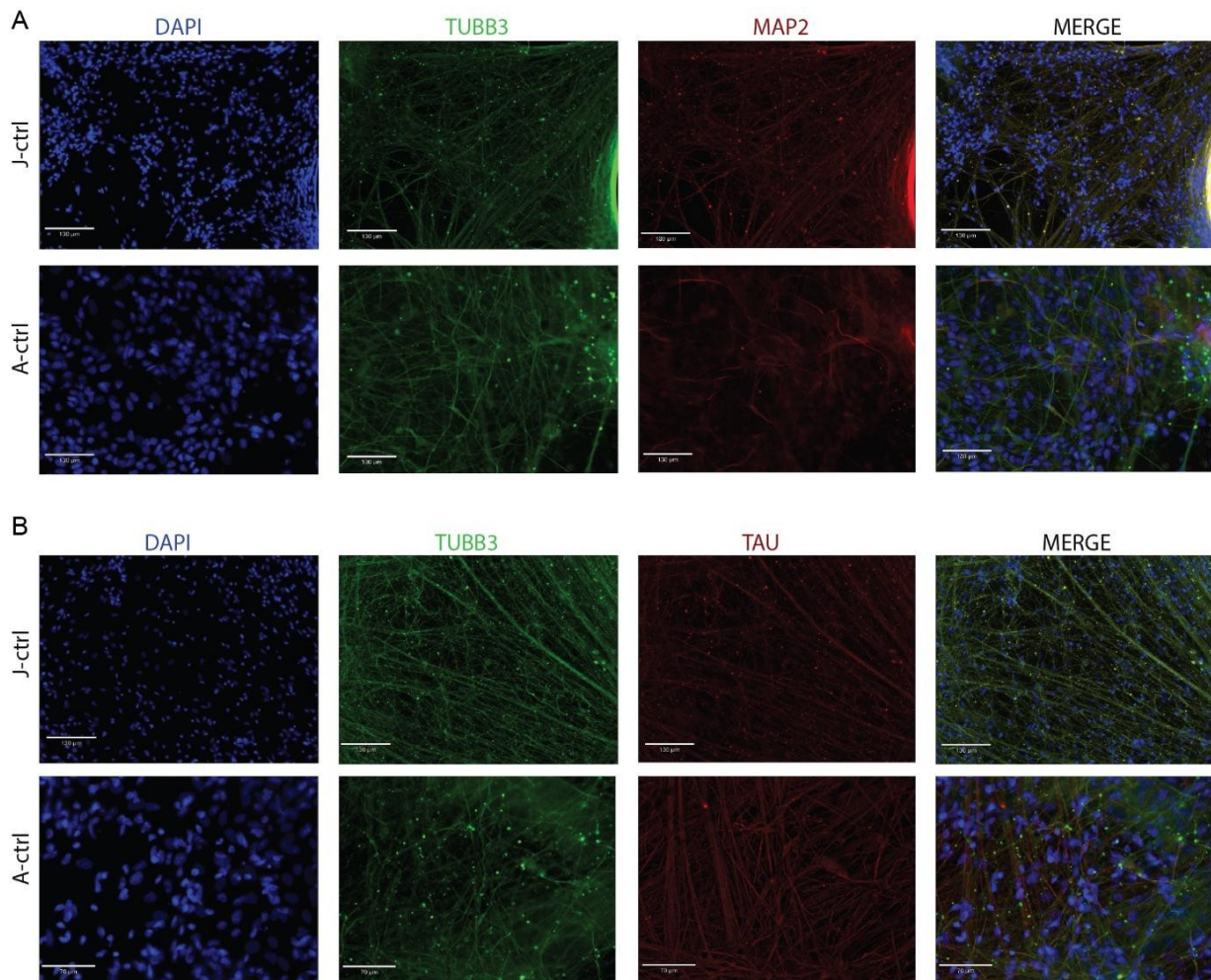


Figure 16. Immunostaining confirmed presence of neurons. Immunostainings were performed for the neuronal marker TUBB3 in combination with MAP2 (A) or TAU (B). Results showed the co-localization of proteins in axonal-like structures. Scale bar: 130 μm , B) A-ctrl scale bar: 70 μm .

4.2 2. Escape gene expression during neuronal differentiation

Bulk RNA-Sequencing technology was used to investigate the expression of X-linked genes from the inactive X-chromosome. Total RNA from iPSCs, NPCs, and neurons derived from 16/98 M-ctrl, 1262/16 J-ctrl, and 1179/17 A-ctrl fibroblasts were sequenced and further analysed with the help of Dr. Stefan Diederich and Dewi Hartwisch, in collaboration with Prof. Susanne Gerber and Dr. Hristo Todorov. To investigate if the sequenced samples had similar cell identities, principal component analyses (PCA) were performed by examining the expression of the top 500 most variably expressed genes. The PCA plot showed clustering of samples corresponding to their cell type (Figure 17.A, B). J-ctrl and M-ctrl iPSCs, as well as J-ctrl and M-ctrl NPCs, clustered together, indicating similar expression patterns within different sequenced samples and different cell lines. J-ctrl and M-ctrl neurons formed two separate clusters, confirming consistent gene expression pattern within neuronal samples in the cell lines and indicating differences in gene expression between the two different cell lines (Figure 17.A). A separate PCA plot for A-ctrl RNA-sequencing was generated due to different library preparation and sequencing conditions (see section 4.2.9.1, Figure 17.B). Sequenced samples clustered together according to cell type, confirming that the samples had similar gene expression profiles.

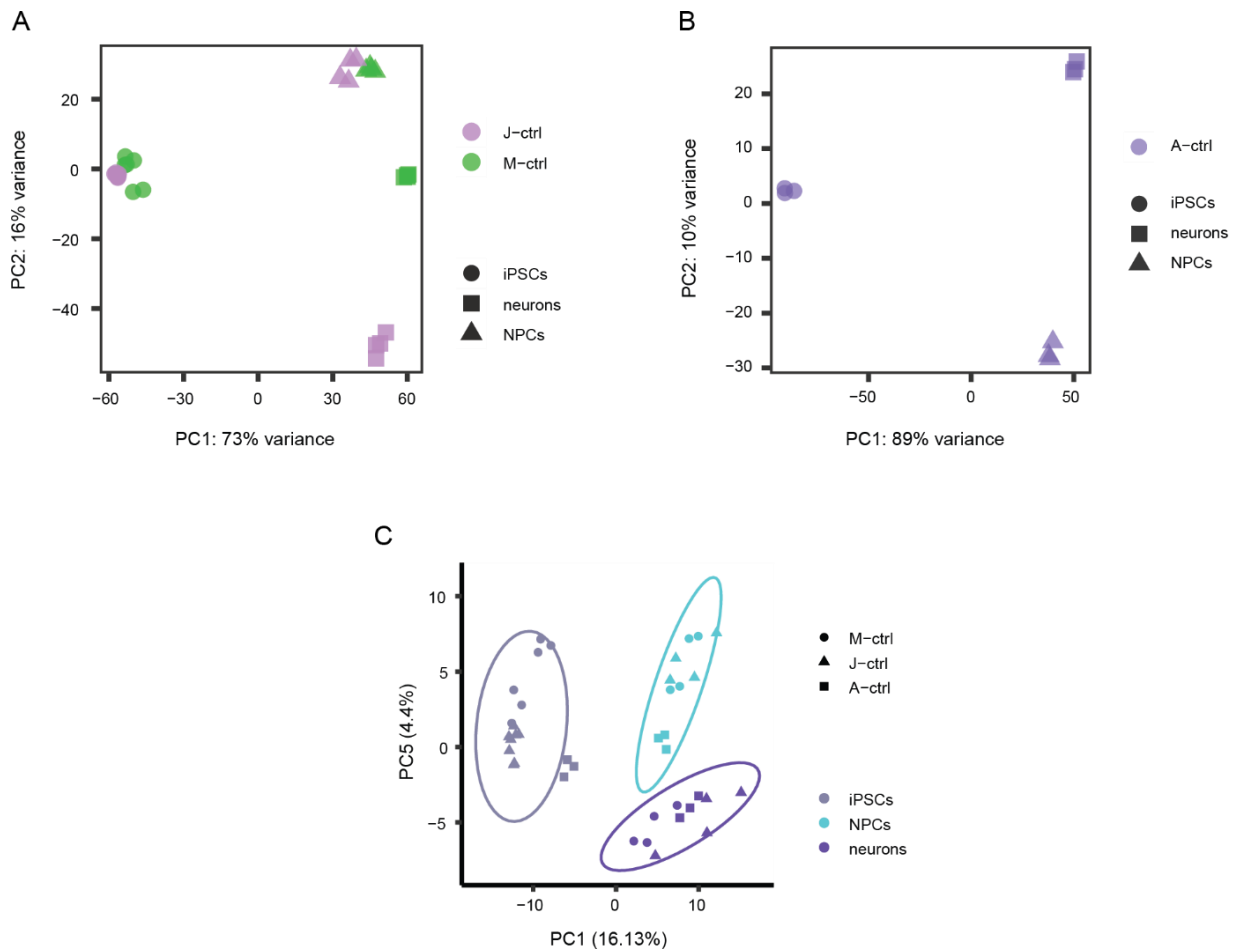


Figure 17. PCA plots confirmed the cell identity of sequenced samples. Sequenced samples clustered depending on the cell type for J-ctrl and M-ctrl (A), and A-ctrl (B). n=6 for M-ctrl iPSCs, n=4 for M-ctrl NPCs and neurons. n=4 for J-ctrl iPSCs, NPCs, and neurons. n=3 for A-ctrl iPSCs, NPCs, and neurons. C) iPSCs, NPCs, and neurons clustered within the 95% confidence ellipses based on their allele-specific expression patterns of X-linked genomic variants.

4.2.1 Dynamic expression from the inactive X-chromosome

Due to the clonal status of the iPSCs, temporal tracking of inactive X-chromosome gene expression during neuronal differentiation was possible. Using iPSC samples as the starting point, expression from the inactive X-chromosome was monitored throughout *in vitro* differentiation of iPSCs into neuronal precursor cells (NPCs) and post-mitotic neurons. Expression of X-linked heterozygous SNPs, insertions, and deletions were used for a global characterization of the allelic expression of X-linked genes from RNA-seq data, enabling the allele-specific resolution analysis of X-linked gene expression. All variants mapping to X-linked genes were taken into consideration. To ensure the expression of the variants, a cut-off of at least five reads mapping to each variant per sample was applied. For each variant, the reference allele was defined as the allele with the majority of mapped reads. The allelic ratio was then calculated based on the ratio between the reference allele and the total number of reads covering the variant site. Based on the variant calling on the X-chromosome, principal component analysis (PCA) was conducted. The PCA plot demonstrated that iPSC, NPC, and neuron samples from the three different

cell lines exhibited similar allelic expression patterns of X-linked genes according to cell type, mostly falling within the 95% confidence ellipses for each cell type (Figure 17.C).

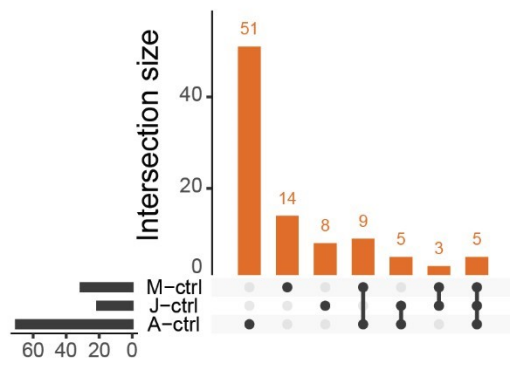
Thanks to allele-specific variant calling and taking into account that iPSCs exhibit skewed X-chromosome inactivation, with always one X-chromosome active and one inactive in all the cells, we were able to trace gene expression from both the active and the inactive X-chromosome. Surprisingly, a new category of facultative escapee genes, which are expressed from the inactive X-chromosome only during neural differentiation, were discovered. These so-called reactivated genes were identified based on variants with an allelic ratio < 0.025 or > 0.975 in iPSCs, indicating monoallelic expression only from the active X-chromosome, and an allelic ratio greater than 0.025 or lower than 0.975 in NPCs and/or neurons, indicating biallelic expression of the same variant site in NPCs and/or neurons. This data was then transformed into probability of expression, and only variants with adjusted P values < 0.01 were considered (see section 4.2.9.3).

Constitutive escapee genes were identified as genes with biallelic expression in iPSCs, NPCs, and neurons. Variants with an allelic ratio greater than 0.1 in all three cell types were considered biallelically expressed and assigned to the full-escapee category. A third and new category of facultative escapee genes, termed late-silenced genes, was identified. These genes exhibited biallelic expression in iPSCs and monoallelic expression in NPCs and neurons. Late-silenced genes were identified by listing all variants in NPCs and neurons with an allelic ratio < 0.025 or > 0.975 , and classifying the same variants with allelic ratio > 0.025 or < 0.975 in iPSCs as late-silenced variants.

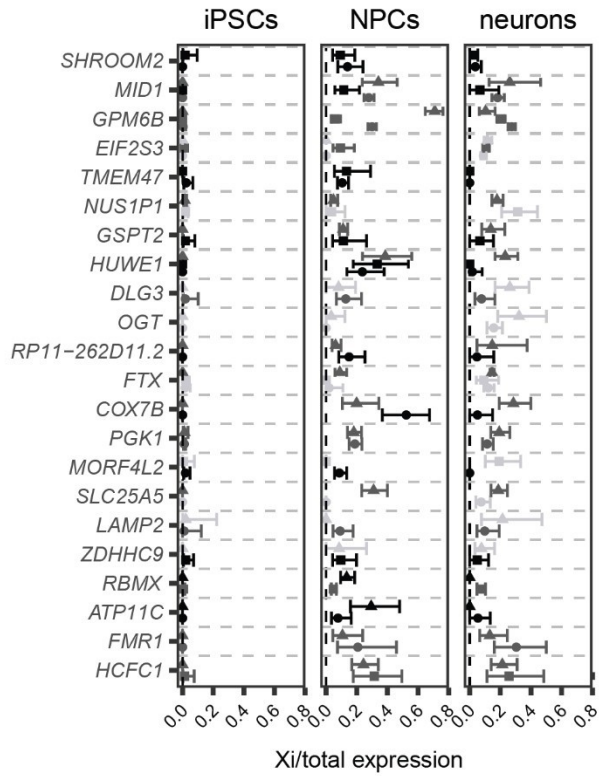
X-linked genes were then assigned to one of the categories (reactivated, full-escapees, or late-silenced genes) based on the heterozygously expressed variants contained in the gene (see section 4.2.9.4). Genes with more than one variant belonging to different categories were manually assigned to a single category based on a set of rules and in a blinded manner (see section 4.2.9.4). Surprisingly, we discovered that a total of 95 X-linked genes were reactivated genes, monoallelically expressed in iPSCs and biallelically expressed in NPCs and/or neurons. Twenty-two of those genes were reactivated in at least two cell lines, with five of those genes reactivated in all three cell lines (Figure 18.A, B). We also identified a total of 81 full-escapee genes that were biallelically expressed in iPSCs, NPCs, and neurons. Thirty-five genes were detected as full-escapees in more than one cell line, and sixteen of them were annotated as full-escapee in all three cell lines (Figure 18.C, D). Additionally, we identified genes assigned to the late-silenced category, a new category of facultative escapee genes, which were expressed from both the active and inactive X-chromosomes in iPSCs and monoallelically expressed from the active X-chromosome in NPCs and neurons (Figure 18.E, F). A total of 29 genes were classified as late-silenced genes, with five genes listed as late-silenced in at least two cell lines.

Taken together, these results demonstrate a dynamic expression from the inactive X-chromosome during in vitro neural differentiation in humans, and identify two novel categories of facultative escapee genes: reactivated genes and late-silenced genes.

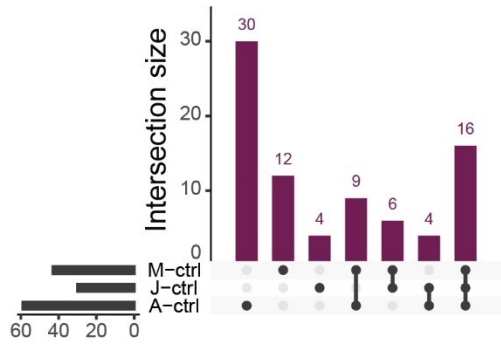
A Reactivated genes



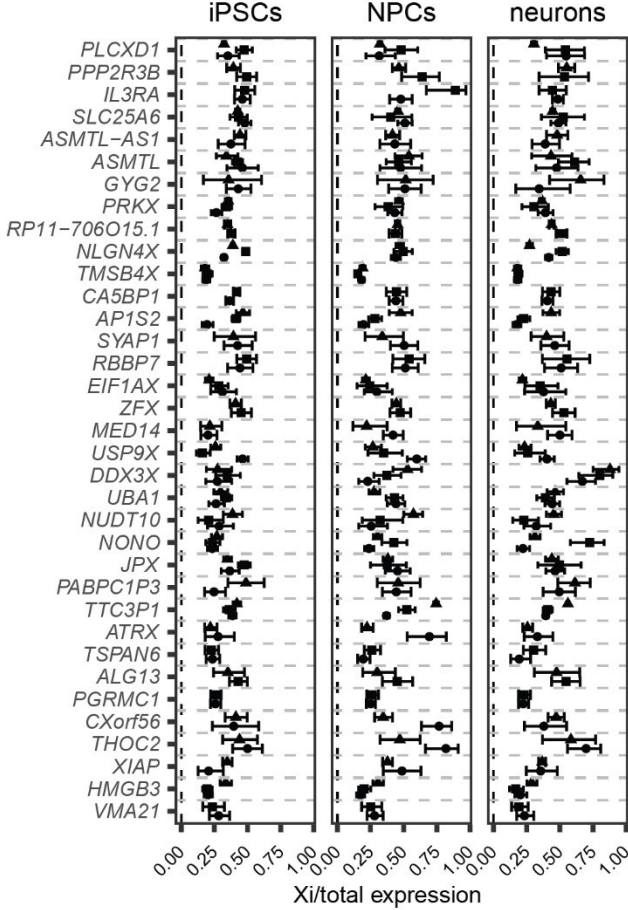
B Reactivated genes



C Full-escapee genes



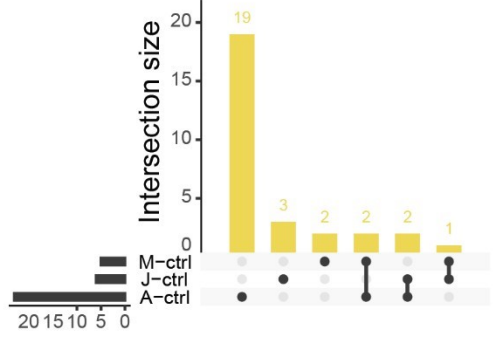
D Full-escapee genes



Reactivated in

- M-ctrl
- J-ctrl
- ▲ A-ctrl
- NPCs
- neurons
- NPCs & neurons

E Late-silenced genes



F Late-silenced genes

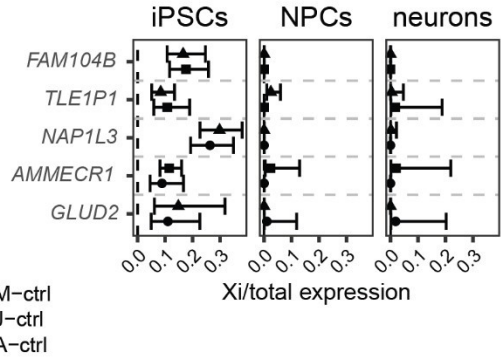


Figure 18. Dynamic X-chromosomal expression during neural differentiation of iPSCs. A-C-E) Bar plot showing the number of reactivated, full-escapee, and late-silenced genes detected for each single cell line and in more than a cell line. B-D-F) Scatter plots showing the expression ratio of reactivated, full-escapee and late-silenced genes from the inactive X-chromosome compared to total allelic mRNA expression. Only genes detected in more than a cell line are reported. Reactivated genes are biallelically expressed in NPCs and/or neurons but not in iPSCs (A, B). Full-escapee genes showed similar biallelic expression in all three cell types (C, D). For late-silenced genes biallelic expression is detected only in iPSCs and not in NPCs and neurons (E, F). n=6 for M-ctrl iPSCs, n=4 for M-ctrl NPCs and neurons. n=4 for J-ctrl iPSCs, NPCs, and neurons. n=3 for A-ctrl iPSCs, NPCs, and neurons.

In humans, facultative and constitutive escape genes form clusters on the X-chromosome corresponding to specific Topologically Associated Domain (TAD)-like domains, characterized by open chromatin regions and thus active gene expression^{447,463,464}. To investigate whether the facultative and constitutive escapee genes clustered in specific regions of the X-chromosome, we highlighted the cytogenetic positions of all detected full-escapees (reported in purple), reactivated (orange), and late-silenced (yellow) escape genes (Figure 19.A). Genes annotated with names are reported in more than one cell line. A density plot was used to understand if these genes were clustering together on the X-chromosome (Figure 19.B). Based on the gene distribution along the X-chromosome, reactivated genes were the only group localizing closer to each other than expected by chance, as indicated by the distance of the peak of the orange curve to the median pairwise distance (orange line) and by the p value < 0.05. Taken together, these data suggest that the identified reactivated genes form clusters along the X-chromosome, likely associated with chromatin regions that exhibit active expression from the inactive X-chromosome in cells during neuronal differentiation.

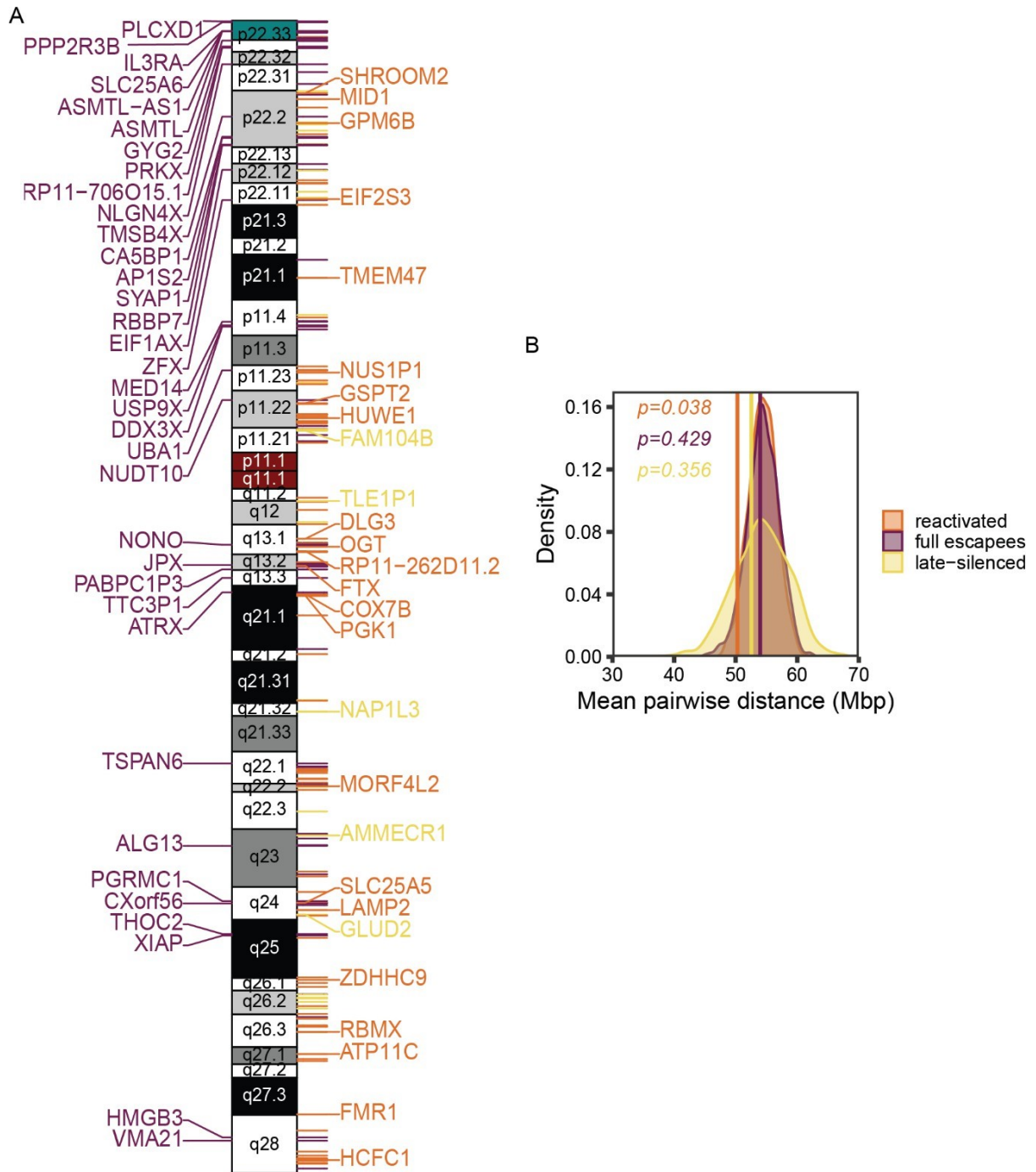


Figure 19. Distribution of facultative and constitutive escapee along the X-chromosome. A) Ideogram showing the cytogenetic bands of the X-chromosome and the location of the full escapee (in purple), reactivated (orange), and late-silenced (yellow) genes. Each line corresponds to a gene listed in one of those categories. Genes labelled with the name correspond to genes detected in more than one cell line. B) Density plot showing the distribution of 1000 X-linked genes randomly selected divided in reactivated (orange), late-silenced (yellow), and full-escapee (purple) group of genes. The curve indicates the variation of the gene distribution in the different categories of facultative and constitutive escapees. Vertical lines indicate the median pairwise distance. P-values were calculated based on the cumulative distribution function of the normal distribution.

Downstream informatic analyses were performed to further characterize the escape genes identified, to elucidate their function, and to possibly understand the mechanisms leading to the flexible gene

expression of the facultative escapee genes (Figure 20). Based on the database of NDD-associated genes⁴⁴², we estimated that X-chromosomal genes are enriched for NDD-associated genes, as reported in the literature³⁶⁴. The X-chromosome has a 2.6-fold higher frequency of NDD-associated genes compared to the expectation value, calculated by randomly selecting the same number of genes present on the X-chromosome (n=836) from the whole genome and estimating the NDD-associated gene enrichment over 10^6 iteration (Figure 20.A). Using the same approach, we estimated whether the three different categories of escape genes- reactivated, full-escapees, and late-silenced- are enriched for NDD-associated genes. For this enrichment analysis, we considered only genes expressed in NPCs and neurons, based on gene expression data from the RNA-seq. Surprisingly, we found that only the reactivated genes are enriched for NDD-associated genes, with almost a 1.5-fold enrichment calculated over 10^6 iterations of randomly picked 95 genes (same number of reactivated genes) out of the 513 X-chromosomal genes expressed in the NPCs and neurons. Additionally, we intersected the list of reactivated genes with genes associated with neurodevelopmental disorders, finding an average 2-fold enrichment for several intellectual disabilities and brain developmental disorders (Figure 20.E). These results suggest that reactivated genes may play a crucial role in human neuronal development. In further functional downstream analysis I have therefore focused on the reactivated category.

To gain insight into the possible reactivation mechanisms, we investigated the epigenetic state of reactivated genes reported in the literature. We merged our list of reactivated genes with the 15-state chromatin model generated from the Roadmap Epigenomics consortium⁴⁶¹, specifically using data from the fetal male and female brains (Figure 20.B, see section 4.2.9.10). The set of reactivated genes was analyzed for enrichment of active or repressive chromatin marks, and the chromatin states were compared to the rest of X-chromosomal genes expressed in NPCs and neurons, which were not reported as biallelic expressed in literature and in our data. These analyses were performed separately on male and female fetal brain datasets. The bar plot displayed the results of the enrichment analysis (Figure 20.B), indicating that reactivated genes are significantly associated with active transcription states in females (“Flanking active TSS”, “Weak transcription”, “Strong transcription”) but not in males. These results suggest that the epigenetic status of reactivated genes may differ between males and females, reflecting the open chromatin status from both X-chromosomes during fetal brain development in females.

Next, we examined whether the facultative and constitutive escape genes we identified were already reported in the literature as escape genes. Two sets of reported escape genes were used as reference databases: one derived from allele-specific expression analysis in females with skewed X-chromosome inactivation, and another study based on sex-bias expression in 44 human tissues^{301,459} (see section 4.2.9.6). All gene categories showed a significant fold-change enrichment for genes already reported as escapees in the literature. This was calculated based on randomized iterations of the same number of genes of the escape gene categories, randomly picked from all X-linked genes and overlapped with the escape gene list reported in literature (Figure 20.C). Furthermore, we investigated if the biallelic expression observed in our data could be linked to female expression bias in vivo. Based on the sex-specific expression pattern in 44 human tissues reported by Oliva et colleagues, we intersected our categories of inactive and biallelic genes. We found that biallelic genes have a female expression bias in a large number of tissues in vivo, significantly higher than inactive genes (Figure 20.D). By examining the individual categories of escape genes, we found that it is particularly reactivated and fully-escape genes that show biallelic expression in a high number of tissues. This suggests that expression from the inactive X-chromosome significantly contributes to higher female expression in vivo across various tissues, not just in brain tissue.

To further gain insight into the molecular function of the reactivated genes, we generated a protein-protein interaction network based on the reactivated genes and genes only expressed in NPCs and neurons (Figure 20.F). GO-term analyses were included in the plot using a color-coded representation. We found a substantial overrepresentation of terms such as histone/chromatin binding, RNA binding, and protein ubiquitination, essential functions for the biology of the neuronal cells.

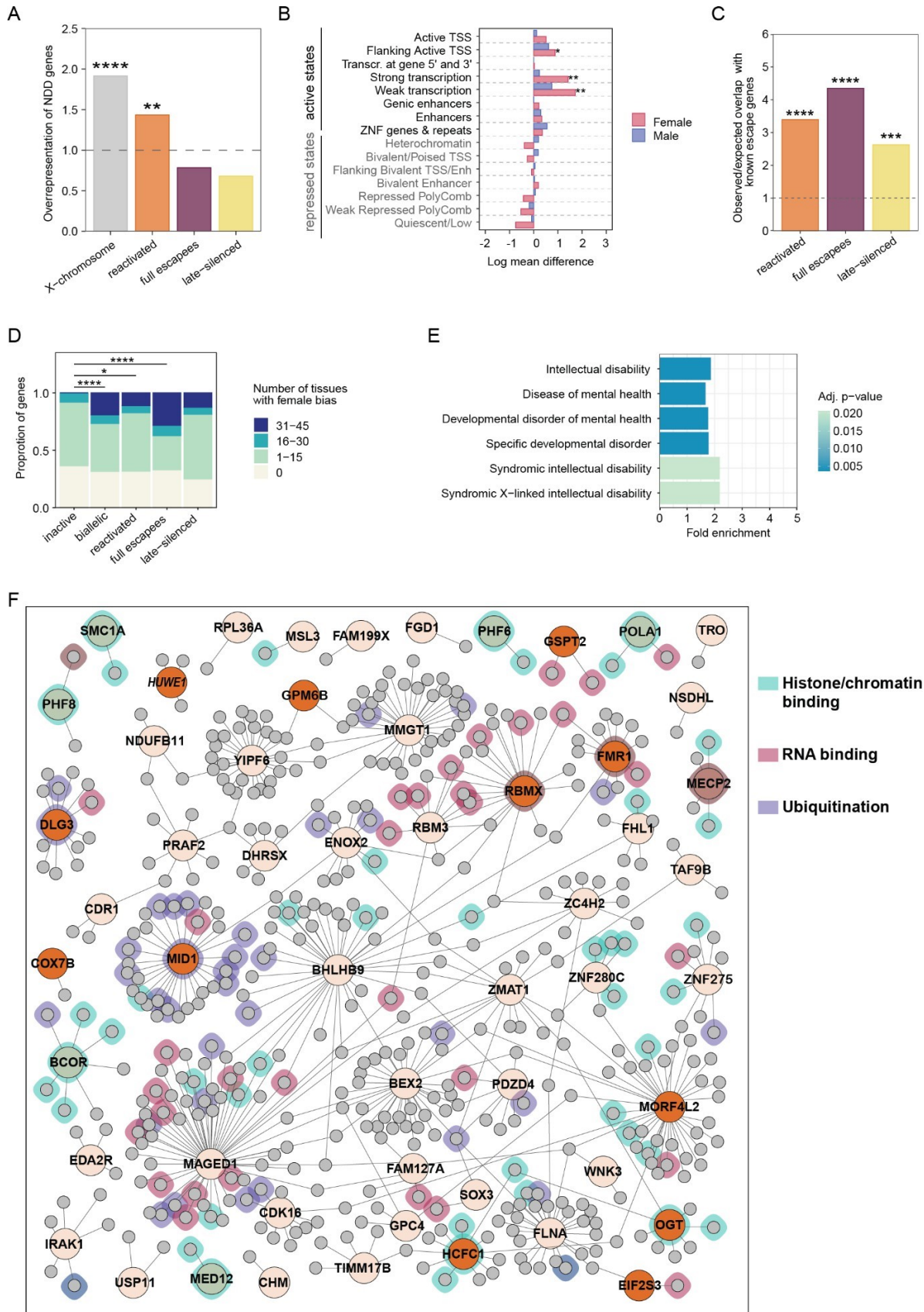


Figure 20. Characterization of in vivo female bias, association with neurodevelopmental disorders, and functional characterization of reactivated genes. A) Barplot showing the overrepresentation of NDD-associated genes on the X-chromosome (1.87-fold change more than expected, $P < 10^{-16}$). Enrichment analyses of X-linked NDD-associated genes expressed in NPCs and neurons in reactivated, full-escapee, and late-silenced genes. Only reactivated category showed a significant enrichment in NDD genes (1.45-fold change compared to expected, $P < 0.01$). B) Barplot representing the fold-change difference between the reactivated gene category and the inactive genes on the X-chromosome. Enrichment analyses were performed in female (pink bars) and male (blue bars). Positive log mean difference represents higher probability that reactivated genes are associated with the chromatin state, and vice versa for negative log mean differences. P values: $P < 0.05$ *, $P < 0.01$ **. C) Barplot highlighting the fold-change of the overrepresentation of reactivated, full-escapee, and late-silenced genes in the escapee genes annotated in human studies. $P < 0.001$ ***, $P < 0.0001$ ****. D) Barplot showing the proportion distribution of the number of tissues with female bias associated with the inactive and biallelic X-linked gene lists, and the different categories of constitutive and facultative escapees. E) Horizontal barplot showing the fold enrichment of genes associated with NDD disorders in the reactivated gene category. F) Protein-protein interaction (PPI) network showing the interaction between reactivated genes and genes only expressed in NPCs and neurons. PPI plot is integrated with GO analysis. Reported overrepresented GO terms are shown as color-coded clouds around the genes. Reactivated genes involved in the GO terms are indicated with the name.

4.2.2 Validation of predicted reactivated genes

Bioinformatic analysis identified a set of constitutive and facultative escape genes. To validate the bioinformatic predictions, we employed several methods such as the Quantification of Allele-Specific Expression by Pyrosequencing (QUASEP) assay. QUASEP assays are based on the pyrosequencing of a heterozygous genomic variant, giving as result the allele-specific relative abundance of the genomic variant. Several genes have been tested by QUASEP in the M-ctrl line previously, confirming the expected pattern of allele-specific expression of the full-escapee *CA5B* gene, the silenced *ZNF185* gene, and the reactivated *MIDI* gene along the neural differentiation trajectory of iPSCs (Thesis, Dr. Stephan Käseberg).

4.2.2.1 QUASEP assay results confirmed reactivation expression pattern of the *GPM6B* gene

The QUASEP assay relies on the presence of expressed heterozygous SNPs or insertions/deletions to detect the allelic expression from the X-chromosomes. To validate allele-specific RNA-seq analysis, heterozygous SNPs detected from the RNA-seq as well as heterozygous SNPs annotated from previous microarray analyses were used. Among the five facultative escape genes reactivated in all three cell lines during neural differentiation, the X-linked *GPM6B* gene was further validated through QUASEP. Two heterozygous SNPs were chosen for QUASEP validation: one SNP identified from the RNA-seq analysis (rs13793, chrX:13,771,230 T>G), and one identified from microarray analysis (rs11095627, chrX:13,772,214 G>A), both located in the 3'-UTR region of the *GPM6B* gene. QUASEP assay performed on genomic DNA confirmed the presence of the heterozygous SNPs in the M and J-ctrl lines, but not in the A-ctrl line (Figure 21).

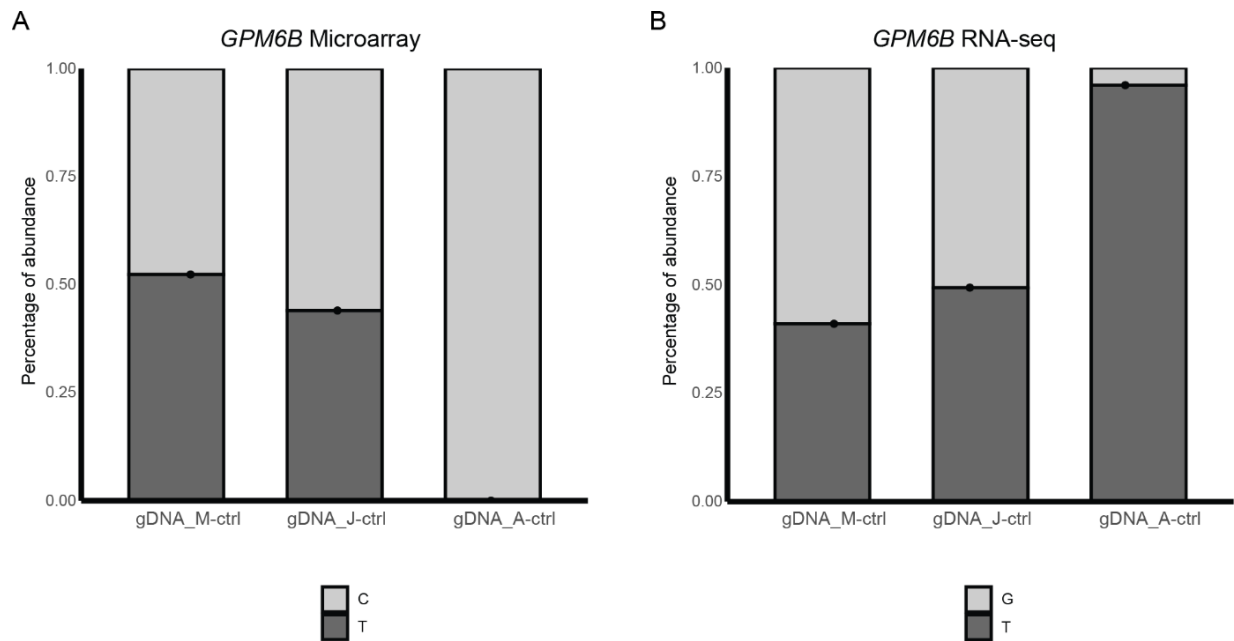


Figure 21. Heterozygous *GPM6B* SNPs presence in M, J, and A-ctrl. Genomic DNA of M, J, and A-ctrl lines were used for QUASEP assay to detect the presence of heterozygous *GPM6B* SNPs. Relative abundance of C and T (A), and G and T (B) indicate heterozygosity of those SNPs in M and J-ctrl, but not in A-ctrl. n=1 for A and B.

Once it was confirmed that the SNPs were present in the J and M-ctrl lines, QUASEP assays were performed to investigate the relative expression of the SNPs in iPSCs, NPCs, and neurons (Figure 22). Genomic DNA was used as a control for the assay. The relative expression of the two heterozygous *GPM6B* SNPs in fibroblast is close to the 50% ratio expected for a mixed population of cells with random X-chromosome inactivation (Figure 22: G \approx 43.7%, T \approx 56.2% for A; and C \approx 57.1%, T \approx 42.8% for B). iPSC samples confirmed strict monoallelic SNP expression, affirming the clonal nature of the iPSC clones (Figure 22: G \approx 1.6%, T \approx 98.3% for A; and C = 100%, T = 0% for B). Relative abundance of both SNPs in NPCs and neurons confirmed the biallelic expression of *GPM6B* detected from the RNA-seq data. NPCs exhibited biallelic SNP expression, indicating that the silenced nucleotide on the inactive X-chromosome is expressed in NPCs (Figure 22: G = 27.375%, T = 72.625% for A; and C \approx 77.583%, T \approx 22.416% for B). Neurons showed a similar relative abundance to NPCs, indicating that biallelic expression persists in post-mitotic neurons (Figure 22: G \approx 24.08%, T \approx 75.92% for A; and C \approx 81.83%, T \approx 18.17% for B). The biallelic expression in NPCs and Neurons was significant compared to iPSCs, confirming the statistically relevant reactivation effect observed in the RNA-seq data.

The heterozygous SNP detected in the RNA-seq analysis was also investigated by Dr. Stephan Käseberg (Thesis, Dr. Stephan Kaseberg) in the M-ctrl line (Figure 22.C). Relative abundance in fibroblasts confirmed the heterogeneous population of cells with random X-chromosome inactivation and subsequent biallelic expression of the SNP (Figure 22.C: G \approx 58.083%, T \approx 41.916%). Both M-ctrl and M-OS-het iPSC clones were used for the QUASEP assay. M-OS-het is an iPSC clone derived from the same iPSC M-line but with the opposite X-chromosome being active compared to M-ctrl. M-ctrl and M-OS-het iPSC samples exhibited a clear monoallelic expression pattern of the SNP in iPSCs compared to fibroblast samples, confirming the clonal nature of the iPSC clones (Figure 22.C: G \approx 99.083%, T \approx 0.916% for M-ctrl and G \approx 11.83%, T \approx 88.16% for M-OS-het). A significantly increased expression from the inactive allele was detected in NPCs (Figure 22.C: G \approx 76.4%, T \approx 23.6% for M-ctrl and G \approx 19.7%, T \approx 80.3% for M-OS-het). Similar relative abundance was detected in neurons (Figure 22.C: G

$\approx 78.3\%$, T $\approx 21.7\%$ for M-ctrl and G = 24.25%, T = 75.75% for M-OS-het). These data confirmed that *GPM6B* is reactivated during neural differentiation of M-ctrl and M-OS-het iPSCs. Furthermore, the extent of reactivation of the SNPs investigated via QUASEP assay between the M and J-lines is similar, around 30%, suggesting that the expression from the inactive X-chromosome is lower compared to the active allele or that not all cells in the mixture reactivate the inactive allele. This finding is consistent with reported literature and RNA-seq data⁴⁶⁵.

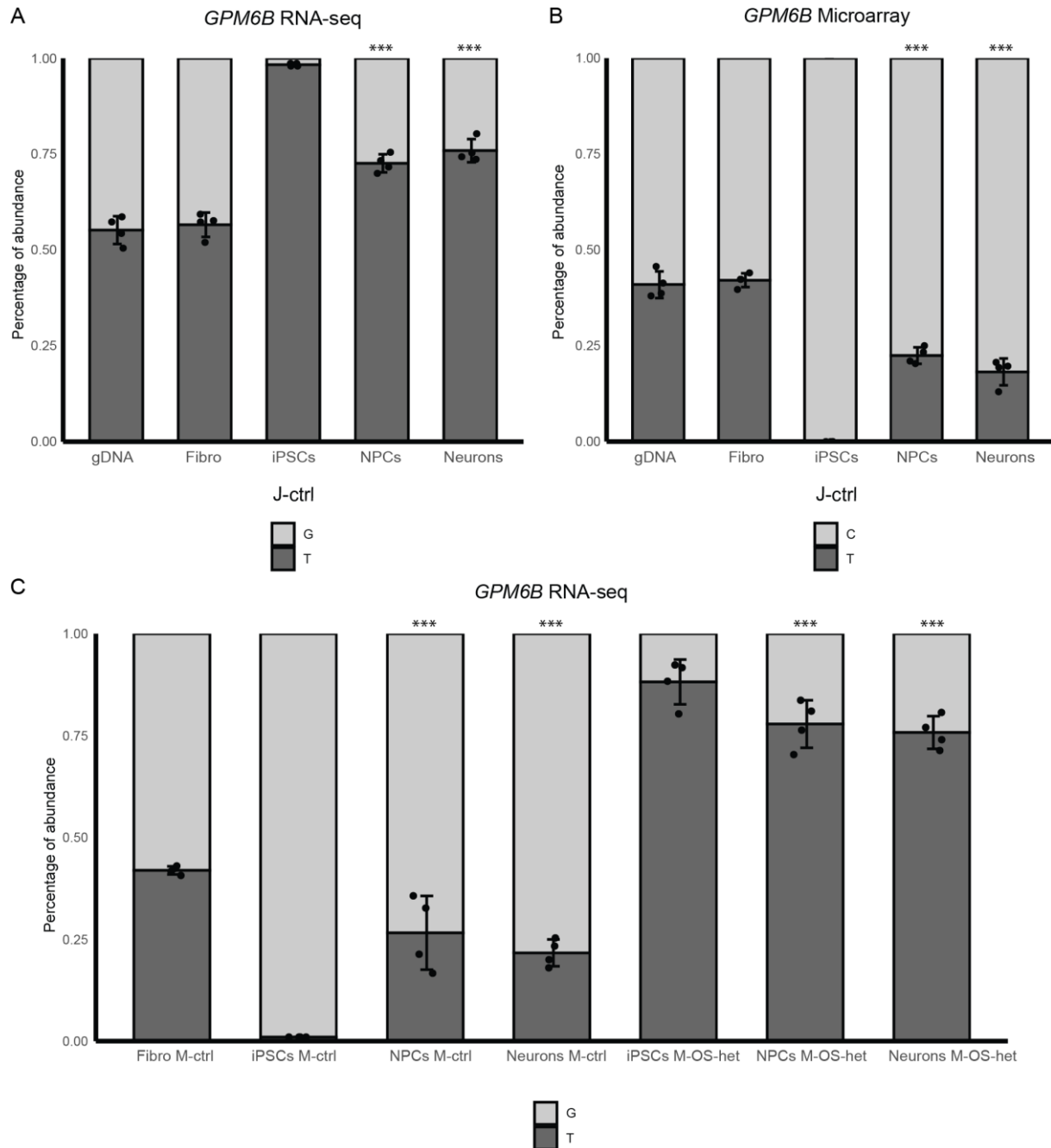


Figure 22. QUASEP analysis confirmed reactivation of *GPM6B* gene in neural cells. *GPM6B* SNPs relative allele-specific abundance in J-ctrl (A-B), M-ctrl and M-OS-het (C). n=4 for all samples. NPCs and neurons were compared with corresponding iPSCs by using One-way ANOVA followed by Tukey's multiple comparison test. Significance is assigned based on the P values: P<0.05 *, P<0.01 **, P<0.001 ***, and P>0.05 non-significant.

4.2.2.2 Allele-specific RT-PCR and QUASEP assay confirmed *MIDI* reactivation

The *MIDI* gene is one of the five genes detected as reactivated in all three cell lines. To confirm *MIDI* reactivation, a QUASEP assay was designed for a heterozygous SNP detected in the RNA-seq data. The heterozygous SNP was identified in the A-ctrl line (rs5934908, chrX: 10,435,519 T>C). Genomic DNA analysis confirmed the presence of the heterozygous SNP in the A-ctrl line (Figure 23). The biallelic expression of the *MIDI* SNP in fibroblasts confirmed a heterogeneous fibroblast population with random X-chromosome inactivation (Figure 23: T = 28.4%, C = 71.6%). In iPSCs, the SNP exhibited monoallelic expression, confirming the clonal status of the iPSC clone (Figure 23: T = 9.6%, C = 90.4%). In NPCs and neurons, the heterozygous SNP showed biallelic expression, confirming the reactivation of *MIDI* in neural cells (Figure 23: T = 37.2%, C = 62.8% for NPCs; T = 48.7%, C = 51.3% for neurons). This data supports the RNA-seq findings, demonstrating that *MIDI* is reactivated during neural differentiation in the A-ctrl line.

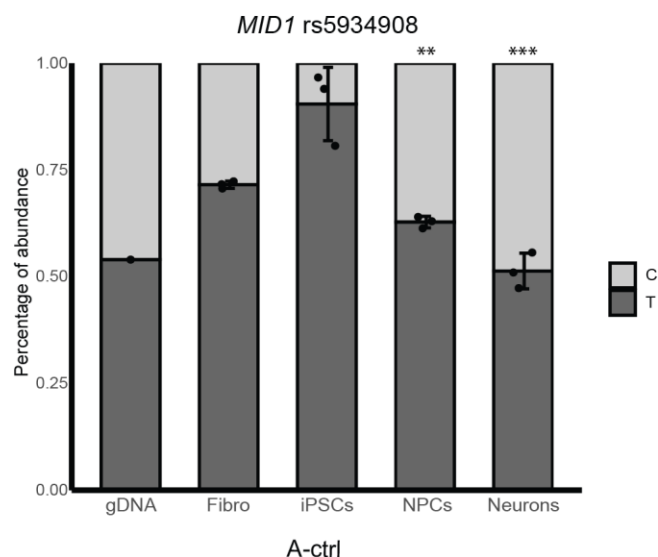


Figure 23. QUASEP analysis confirmed reactivation of *MIDI* in A-ctrl neural cells. *MIDI* SNP allele-specific expression in A-ctrl. QUASEP results confirmed the reactivation of the SNP found in the RNA-seq data. n=1 for gDNA, n=3 for fibroblasts, iPSCs, NPCs, and neurons. NPCs and neurons were compared with corresponding iPSCs and by using One-way ANOVA followed by Tukey's multiple comparison test. Significance is assigned based on the P values: P<0.01 **, P<0.001 ***, and P>0.05 non-significant.

The M-ctrl cell line is derived from a patient heterozygous for a *MIDI* variant. To detect the 4 bp deletion, allele-specific RT-PCR and QUASEP assay were developed and previously performed in iPSCs, NPCs, and neurons (Thesis, Dr. Stephan Käseberg). Both methods demonstrated clear monoallelic expression in iPSCs and concomitant reactivation of the *MIDI* allele located on the inactive X-chromosome during neural differentiation.

To further investigate the role of the 4 bp *MIDI* variant, CRISPR-Cas9 was employed to introduce the same mutation into J-ctrl and A-ctrl lines. CRISPR-edited iPSC clones were screened for the *MIDI* variant, and heterozygous as well as homozygous clones were selected via Sanger sequencing and the genotype confirmed by using allele-specific RT-PCR for *MIDI* (Figure 24). CRISPR-edited clones were

single-cell sorted after electroporation, and clonal heterozygous clones were selected based on the expression of the mutated *MIDI* allele. J-OS het clones were successfully generated by Dr. Stephan Käseberg, while the A-OS het clones were generated by Azza Soliman, Hala Ibrahim, and Verena Engelhardt.

Allele-specific RT-PCR confirmed the clonal nature of the J-OS het and A-OS het clones, as well as the expression of the variant *MIDI* allele carrying the 4 bp deletion (Figure 24). Fibroblasts derived from 1262/16 control and J-ctrl iPSCs were used as control for the allele-specific RT-PCR. Both fibroblasts and J-ctrl iPSCs showed only the expression of the wild type *MIDI* allele, confirming the absence of the mutant variant in the healthy donor (Figure 24.A). The J-OS het clone displayed a strict monoallelic expression of the mutated *MIDI* isoform, confirming the presence of the mutant variant heterozygously and the skewed X-chromosomal status of the iPS clone (Figure 24.A). The J-OS homozygous clone confirmed the absence of the wild type *MIDI* allele by presenting the monoallelic expression of the mutant *MIDI* transcript (Figure 24.A). Same results were observed for the allele-specific RT-PCR with the A-OS het clone. M-ctrl and M-OS het iPSC samples were used as controls for the PCR reaction, while A-ctrl iPSCs were used to confirm that the A-ctrl cell line does not exhibit the *MIDI* variant (Figure 24.B). The A-OS heterozygous iPSCs showed a strict monoallelic expression of the mutated *MIDI* allele, confirming the presence of the *MIDI* 4 bp deletion and the clonal status of the iPSCs (Figure 24.B). Homozygous clones showed the expression of only the mutant *MIDI* allele, confirming the absence of the wild-type *MIDI* transcript (Figure 24.B).

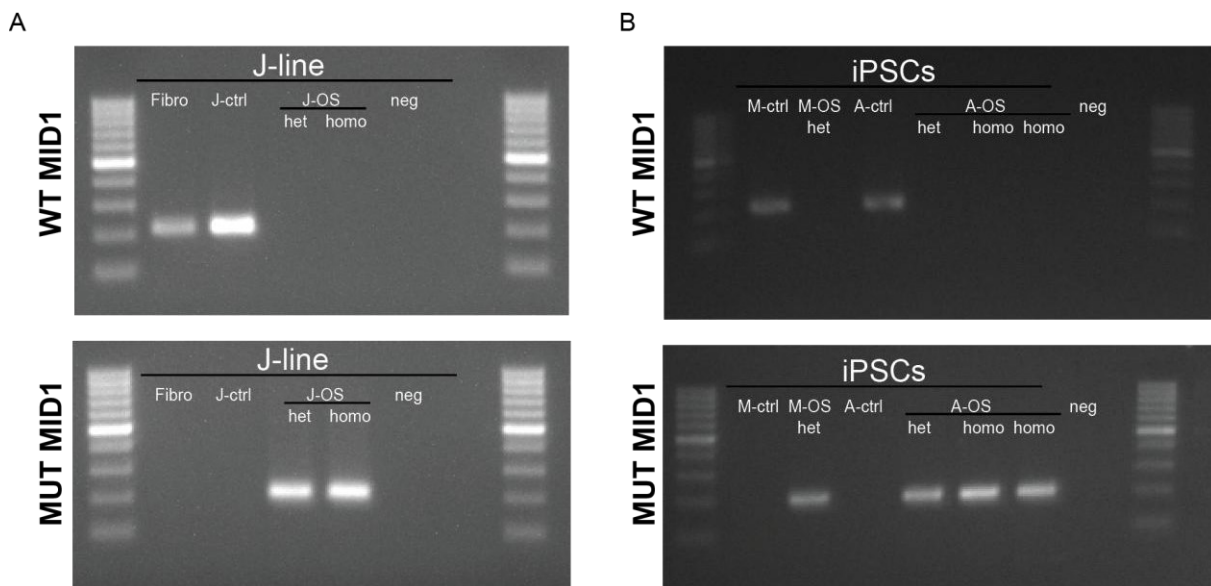


Figure 24. Allele-specific RT-PCR confirmed the presence of *MIDI* patient-specific mutation in CRISPR-edited J and A iPS clones. CRISPR-edited J and A iPSCs were screened for heterozygous and homozygous presence of *MIDI* 4bp deletion mutation. The presence of the mutation was further confirmed by using allele-specific RT-PCR for *MIDI* wild type and mutant transcripts. Selected clones J-OS het and homo (A), as well as A-OS het and homo (B), confirmed the monoallelic expression of *MIDI* transcript. A) Fibroblasts and J-ctrl were used as control. B) M-ctrl, M-OS, and A-ctrl were used as control.

J-OS heterozygous iPSCs were successfully differentiated into NPCs and neurons following the established workflow. Immunostaining and RT-qPCR results confirmed the presence of NPC-specific markers and the downregulation of iPSC-specific markers (Figure 25).

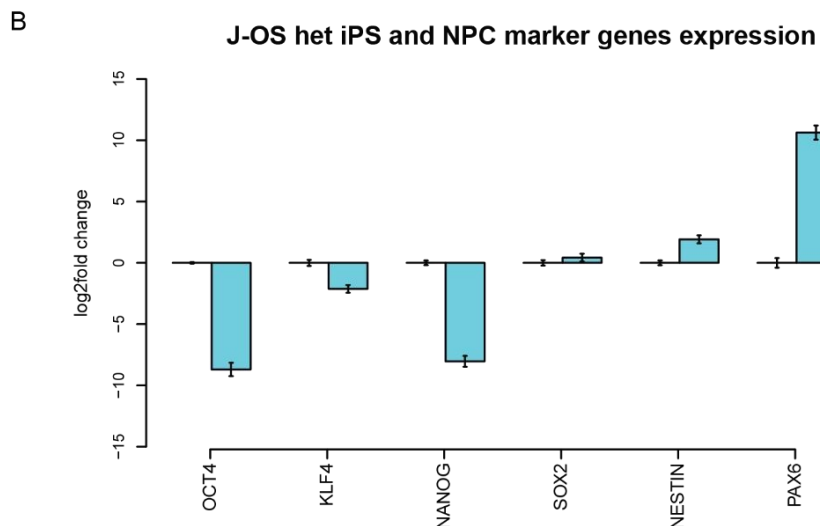
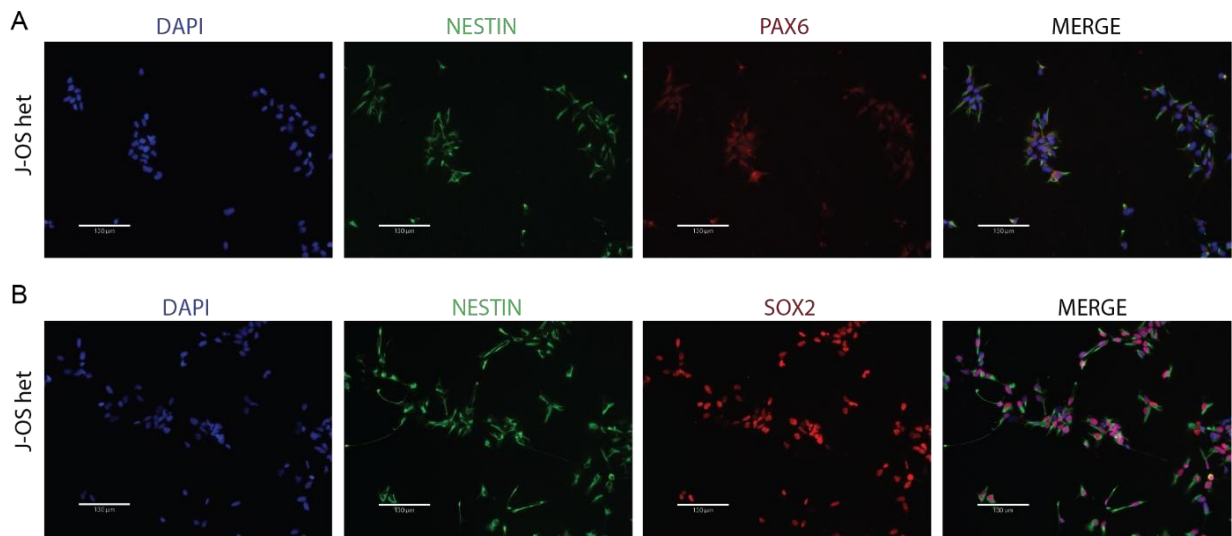


Figure 25. Characterization of successful differentiation of J-OS het NPCs. CRISPR-edited J-OS heterozygous iPSCs were differentiated into NPCs by following the previously established workflow. A) Immunostaining analysis confirmed the presence of the protein markers PAX6, SOX2, and NESTIN. Scale bar = 130 μ m. B) RT-qPCR analysis of iPSC (*OCT4*, *KLF4*, and *NANOG*) and NPC (*SOX2*, *NESTIN*, and *PAX6*) marker genes. n=4 for J-OS het iPSCs and NPCs.

MIDI reactivation was investigated by tracing the allele-specific expression of the wild type and variant alleles via allele-specific RT-PCR and QUASEP assay during 2D neural differentiation (Figure 26). Allele-specific RT-PCR showed that while iPSCs are characterized by monoallelic *MIDI* expression, *MIDI* transcripts are expressed biallelically in NPCs, confirming the expression of the inactive *MIDI* allele in these cells (Figure 26.A). Furthermore, the expression of the inactive allele increased with successive cell passages (number of times cells were re-plated in a new plate), as indicated by the intensity of the *MIDI* band (Figure 26.A). A similar trend was observed in the QUASEP assay results for the *MIDI* 4-bp deletion variant (Figure 26.B). J-OS het NPCs exhibited increased biallelic expression of *MIDI* correlated with higher cell passages (Figure 26.B). The extent of the biallelic expression reached a plateau of approximately 15% expression from the inactive *MIDI* allele around passage 12,

and this level of biallelic expression was maintained in neurons (Figure 26.C). Taken together, these results confirmed *MIDI* reactivation in the J-OS het cell line and suggest that the extent of *MIDI* reactivation increases during neural differentiation.

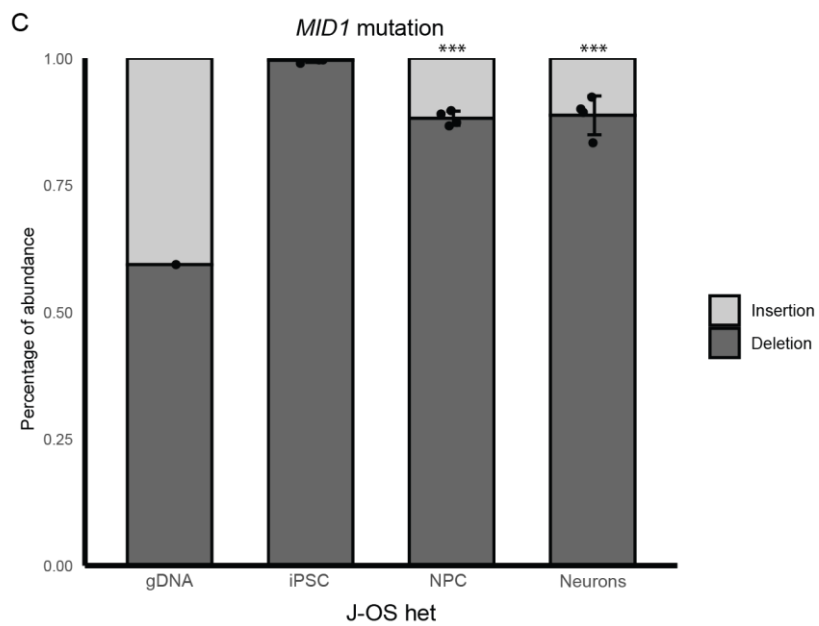
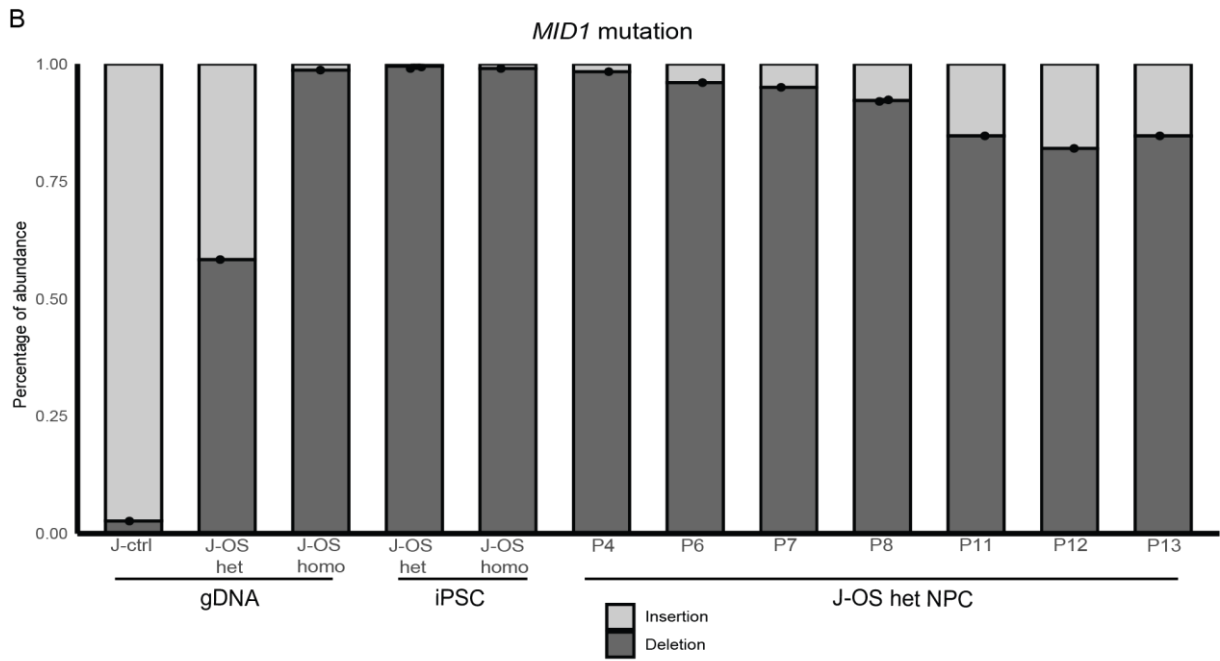
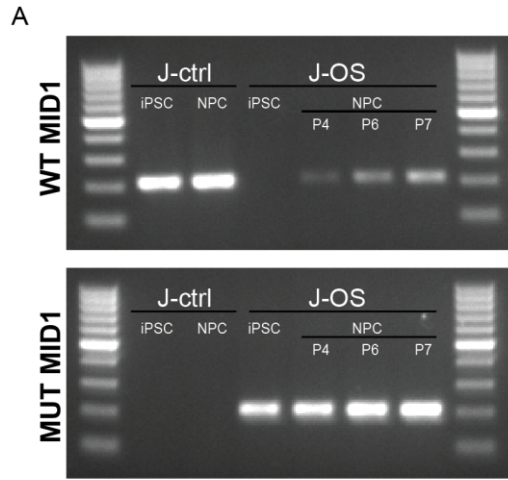


Figure 26. J-OS het cells express *MIDI* inactive allele during neural differentiation. A) Allele-specific RT-PCR for *MIDI* wild type and mutant allele in J-ctrl and J-OS het iPSCs and NPCs. P indicates the passage number of the sample tested. B) QUASEP assay results for the allelic expression of *MIDI* mutation in J-OS het iPSCs and NPCs. C) QUASEP results for *MIDI* mutation in J-OS het iPSCs, NPCs, and neurons. n=1 for gDNA, n=4 for iPSCs, NPCs, and neurons. NPCs and neurons were compared with iPSCs by using One-way ANOVA followed by Tukey's multiple comparison test. Significance is assigned based on the P values: P<0.05 *, P<0.01 **, P<0.001 ***, and P>0.05 non-significant.

An A-OS iPS clone heterozygously carrying the same 4-bp deletion variant in the *MIDI* gene as seen in M- and J-lines was further differentiated into NPCs following the established protocol (Figure 27). Immunostaining analysis confirmed the expression of the NPC-specific marker proteins, as well as the characteristic morphology of the NPCs (Figure 27).

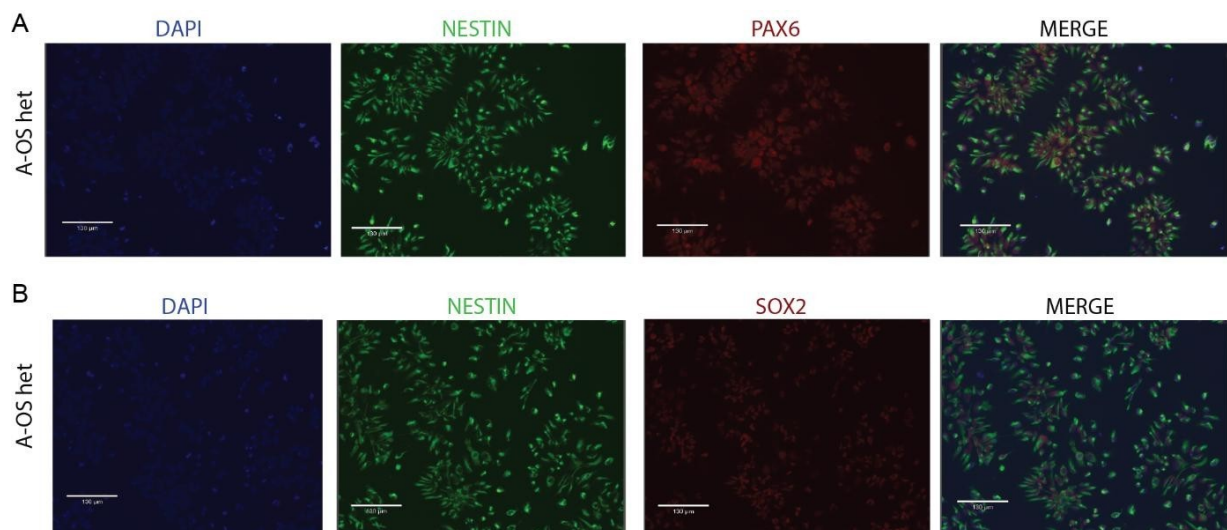


Figure 27. A-OS het NPCs confirmation of successful differentiation. CRISPR-edited A-OS heterozygous iPSCs were differentiated into NPCs by following the previously established workflow. Immunostaining analysis confirmed the presence of the protein markers PAX6, SOX2, and NESTIN. Scale bar = 130 µm.

Allele-specific RT-PCR was used to investigate the dynamic expression of *MIDI* during neural differentiation in the A-OS het clone (Figure 28). M-ctrl iPSC and NPC samples were used as controls for the reaction. iPSCs derived from A-ctrl, A-OS heterozygous and A-OS homozygous lines showed monoallelic *MIDI* expression: the wild-type allele in A-ctrl, and the variant allele in both A-OS het and homo clones. NPCs derived from A-OS heterozygous clone exhibited biallelic expression of *MIDI* transcripts. While the variant allele located on the active X-chromosome maintained stable expression during different passages of the NPCs, the wild type allele showed low expression in the early passages. From passage 4 onward, the expression level of the wild-type allele became similar to that of the mutant allele, confirming the dynamic expression from the inactive X-chromosome.

A

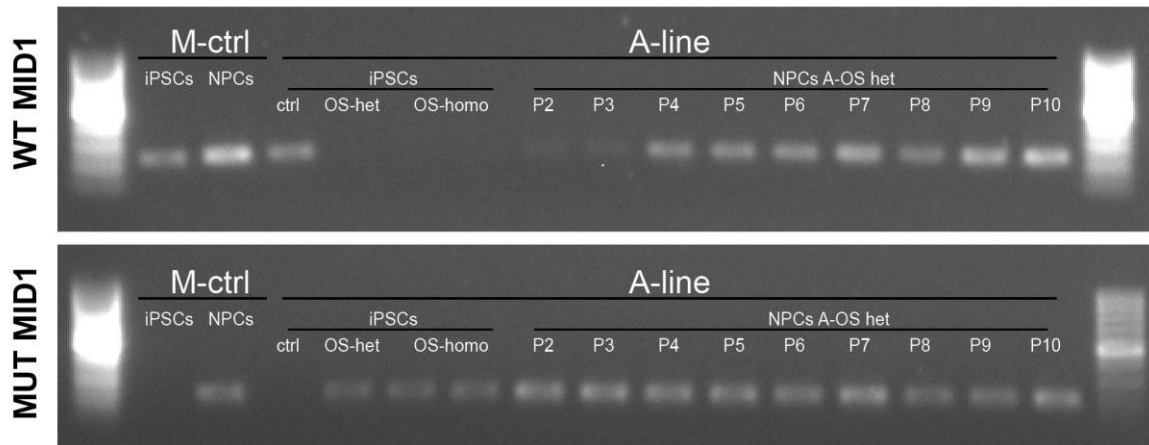


Figure 28. A-OS het cells express *MID1* inactive allele during neural differentiation. Allele-specific RT-PCR for *MID1* wild type and mutant allele in M-ctrl, A-ctrl, A-OS het, and A-OS homo iPSCs and NPCs. P indicates the passage number of the sample tested.

4.3 Erosion of the X-chromosome is not responsible for the dynamic gene expression from the inactive X-chromosome

4.3.1 RNA-seq data revealed heterogeneous presence of eroded and non-eroded cell lines

Expression from the inactive X-chromosome can be altered by the erosion process^{466,467}. Erosion is an artificial mechanism that occurs during the *in vitro* culturing of pluripotent stem cells. It leads to the loss of repressive epigenetic marks and the *XIST* coating of the inactive X-chromosome, resulting in partial loss of X-inactivation and in artificial biallelic expression of X-linked genes. To determine whether erosion of the inactive X-chromosome occurs in the *in vitro* model used in the study and whether this process is responsible for the expression from the inactive X-chromosome, we investigated the status of the inactivation of the X-chromosomes in our cell lines.

Since one of the hallmarks of erosion is the loss of *XIST* expression and coating on the inactive X-chromosome, we employed RNA-seq data to check *XIST* expression (Figure 29.A). *XIST* expression was detected in all three cell lines, but with substantial differences in the degree of expression between them. The highest *XIST* expression was observed in the A-ctrl line, whereas the J-ctrl and M-ctrl lines exhibited low *XIST* expression, consistent with the eroded cell lines reported in the literature. Additionally, we found that the *XIST* expression level is maintained during *in vitro* neural differentiation. This is evidenced by the expression patterns in NPCs and neurons differentiated from the high-*XIST* A-ctrl line and the low-*XIST* M-ctrl and J-ctrl lines.

Eroded cells are characterized by an artificial biallelic expression of X-linked genes, resulting in the expression of most of the X-linked genes being similar to that of autosomal genes rather than being monoallelic. To assess the erosion status of the inactive X-chromosome in the RNA-seq samples, we calculated the ratio of biallelically expressed variants from the X-chromosome and those found in autosomes⁴⁶⁸ (X:A allelic ratio, Figure 29.B). The autosomal ratio was calculated based on the biallelic variants on each autosome, reflecting the average of the autosomal ratios. The autosomal to autosomal allelic ratio mean (A:A allelic ratio) is 1, indicating that genes from autosomes are expressed from both

alleles. The X:A allelic ratio was reported to be lower than 0.39, indicating that the samples were between the inactive (XaXi) and eroded (XaXe) status⁴⁶⁸. Furthermore, the A-line showed the lowest X:A allelic ratio, consistent with its higher *XIST* expression compared to the M- and J-lines. Despite the high X:A allelic ratio, all samples have a significantly lower allelic ratio compared to the autosomal allelic ratio, indicating that the X-linked genes are not entirely biallelic expressed.

Taken together, we identified different levels of *XIST* expression and X:A ratios among the three cell lines used for the allele-specific RNA-seq analyses, indicating that two of our cell lines might exhibit hallmarks of erosion.

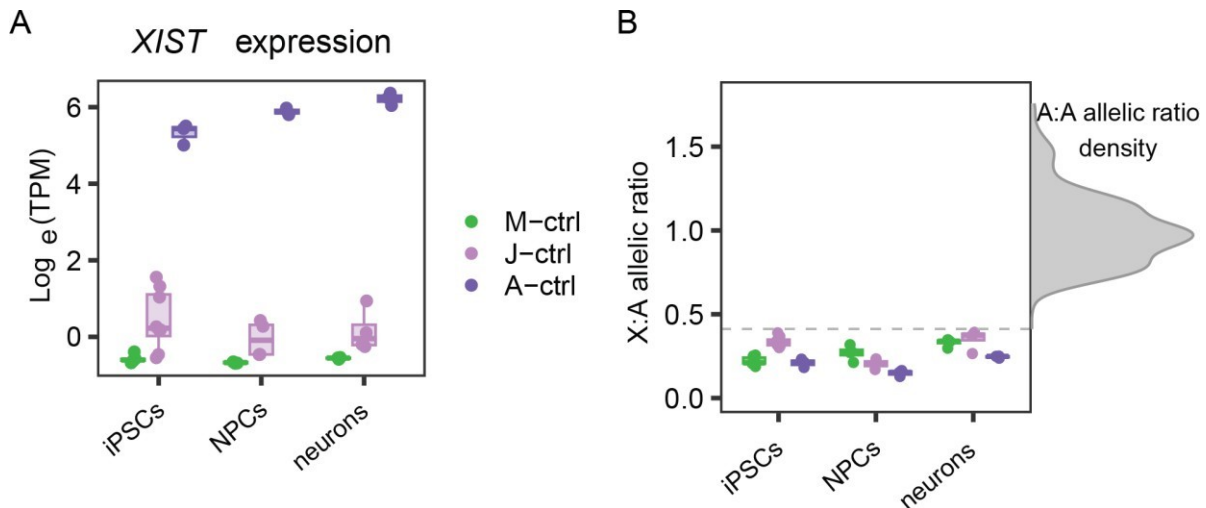


Figure 29. *XIST* expression and X:A allelic ratio showed an heterogeneous population of non-eroded and eroded cell lines. A) *XIST* RNA levels in M-ctrl, J-ctrl, and A-ctrl iPSCs, NPCs, and neurons. *XIST* expression was calculated based on the results of the RNA-sequencing B) Boxplot depicting the X:A allelic ratio showing median, quartiles, and range for the cell types of the specific cell lines. A:A density plot is shown on the right Y-axis of the plot, with the mean value of 1.

4.3.2 Cell line screening for hallmark of erosion revealed complex pattern of erosion in iPSCs and NPCs

The erosion process leads to the artificial loss of repressive histone marks on the inactive X-chromosome. One of the first histone marks to be lost is the repressive H3K27me3 modification, which is found at the promoters of repressed genes and is abundant on the inactive X-chromosome. To further confirm the results from the RNA-seq samples, we performed immunostaining analysis on iPSCs and NPCs from the non-eroded A-line, and the possibly partly eroded M- and J-lines (Figure 30). H3K27me3 accumulates on the inactive X-chromosome, forming a strong nuclear signal in non-eroded cells, representing the inactive X-chromosome territory, as reported in the literature⁴⁴⁶. In eroded cells, the H3K27me3 signal is diffused throughout the nucleus without accumulation on the inactive X-chromosome, indicating the absence of the repressive histone mark on the promoters of X-linked silenced genes. The results of the iPSC staining confirmed that in the high *XIST*-expressing A-ctrl cell line, H3K27me3 accumulated in the nucleus, forming a territory representing the inactive X-chromosome (Figure 30.A). In contrast, in the low *XIST*-expressing M-ctrl and J-ctrl cell lines, few or no cells showed an accumulation of H3K27me3 in the nuclei, but rather a diffuse nuclear signal,

indicating the absence of H3K27me3 accumulation on the inactive X-chromosome and thus a (partial) erosion of the inactive X-chromosome. Similar results were observed for the NPCs (Figure 30.B). High *XIST*-expressing A-ctrl NPCs confirmed the presence of the H3K27me3 territory in the majority of the cells, while NPCs from low *XIST*-expressing M-ctrl and J-ctrl lines showed few or no cells with H3K27me3 accumulation. These results confirmed that erosion is linked to the loss of *XIST* expression, which occurs at the iPSC stage, and that the NPCs inherit the same erosion status as the iPSCs.

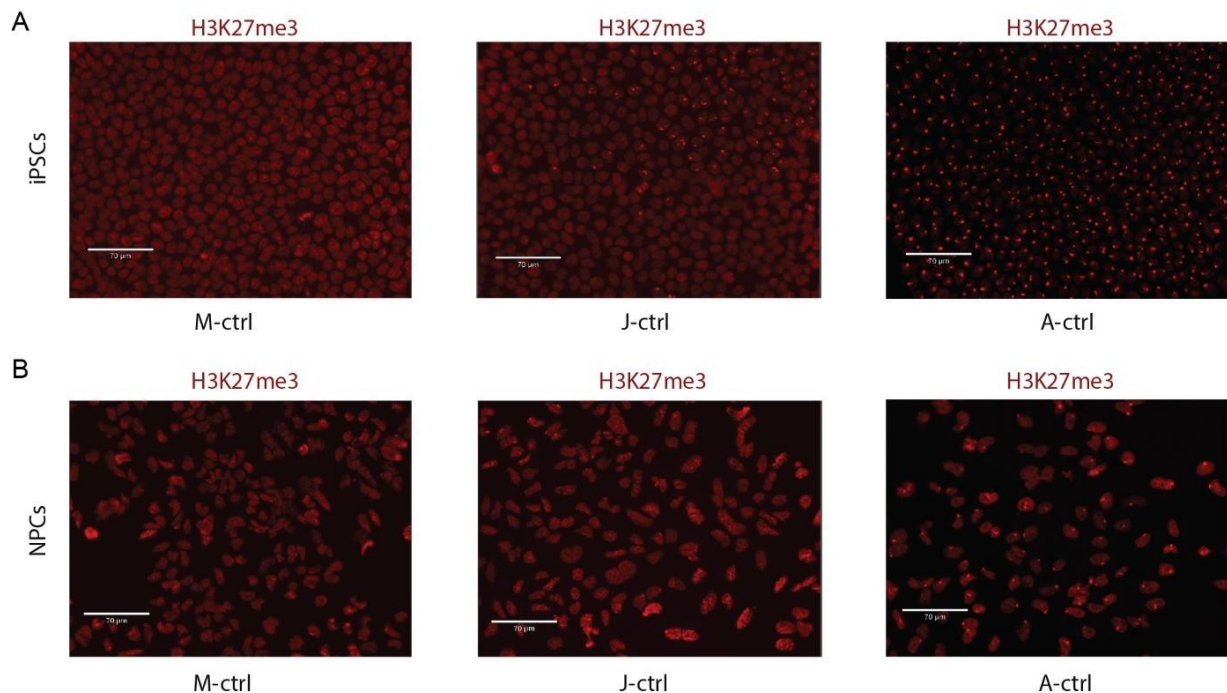


Figure 30. H3K27me3 staining confirmed the erosion status. Immunostaining for H3K27me3 histone modification in iPSCs (A) and NPCs (B). Eroded cells showed a diffuse nuclear signal while non-eroded cells present an accumulation of the signal in the nucleus highlighting the inactive X-chromosome territory. Scale bar = 70 μ m

To further validate the RNA-seq data, we performed *XIST* RT-qPCR analysis on the cell lines used in our experiments (Figure 31). As a control for *XIST* expression, we used RNA from the female fibroblasts from which the iPS lines were derived. Fibroblasts do not undergo erosion and maintain high *XIST* expression while cultured in vitro. For the A-ctrl experiment, fibroblasts were used as a positive control for *XIST* expression, while male iPSCs and eroded iPSCs were used as negative control, lacking *XIST* expression. A-ctrl iPSCs showed similar levels of *XIST* expression compared to the fibroblasts of origin (Figure 31.A). Two independent NPC lines differentiated from A-ctrl iPSCs showed different levels of *XIST* expression. One line, NPC 1, exhibited low *XIST* expression comparable to male and eroded iPS samples, while the other line, NPC 2, had high *XIST* levels comparable to fibroblasts. This data indicates that the NPCs retain the *XIST* expression status of their cells of origin, as reported in the literature⁴⁶⁹. The NPC 2 line was used for the thesis experiment.

For the *XIST* expression analysis in the J-line, male iPSCs lacking *XIST* and non-eroded iPSCs were used as controls (Figure 31.B). *XIST* expression in the J-ctrl iPSCs was lower compared to the fibroblasts of origin and non-eroded iPSCs, indicating the presence of erosion in the iPS samples. However, the expression levels were higher than those in male iPSCs, suggesting that only a portion of the cells were

eroded. CRISPR-edited J-OS heterozygous and homozygous cells showed low *XIST* expression comparable to male iPSCs, suggesting a complete erosion of the cells.

For the M line, non-eroded A-ctrl iPSCs and male iPSCs were used as controls for high and low *XIST* expression, respectively (Figure 31.C). M-ctrl and M-OS het iPSCs showed low *XIST* expression, with levels comparable to those in male iPSCs. This confirmed that the cells were eroded with a complete absence of *XIST* expression. To further understand whether erosion affects the M-line or only the specific clones used in our experiments, two other iPS clones were screened for *XIST* expression: M-ctrl 16, expressing the wild type *MIDI* allele, and M-OS het 19, expressing the mutant *MIDI* allele (Figure 31.D). Fibroblasts were used as high *XIST*-expressing control, while male iPSCs served as negative controls. The expression levels of *XIST* in both iPS clones were lower compared to fibroblasts, indicating that the cells were eroded. However, while M-OS het 19 showed *XIST* expression at levels similar to those in male iPSCs, M-ctrl 16 showed higher *XIST* expression, indicating the presence of non-eroded cells in the samples. Notably, NPCs derived from M-ctrl 16 exhibited *XIST* expression levels comparable to those in fibroblasts, suggesting that the NPCs were not eroded and were likely differentiated from non-eroded cells. M-OS het 19 NPCs showed lower *XIST* expression, but still higher than that in M-OS het 19 iPSCs.

Taken together, monitoring *XIST* expression via qPCR can identify non-eroded and eroded cells. Different cell lines exhibited varying degree of erosion, suggesting that the process is both line-specific and passage specific.

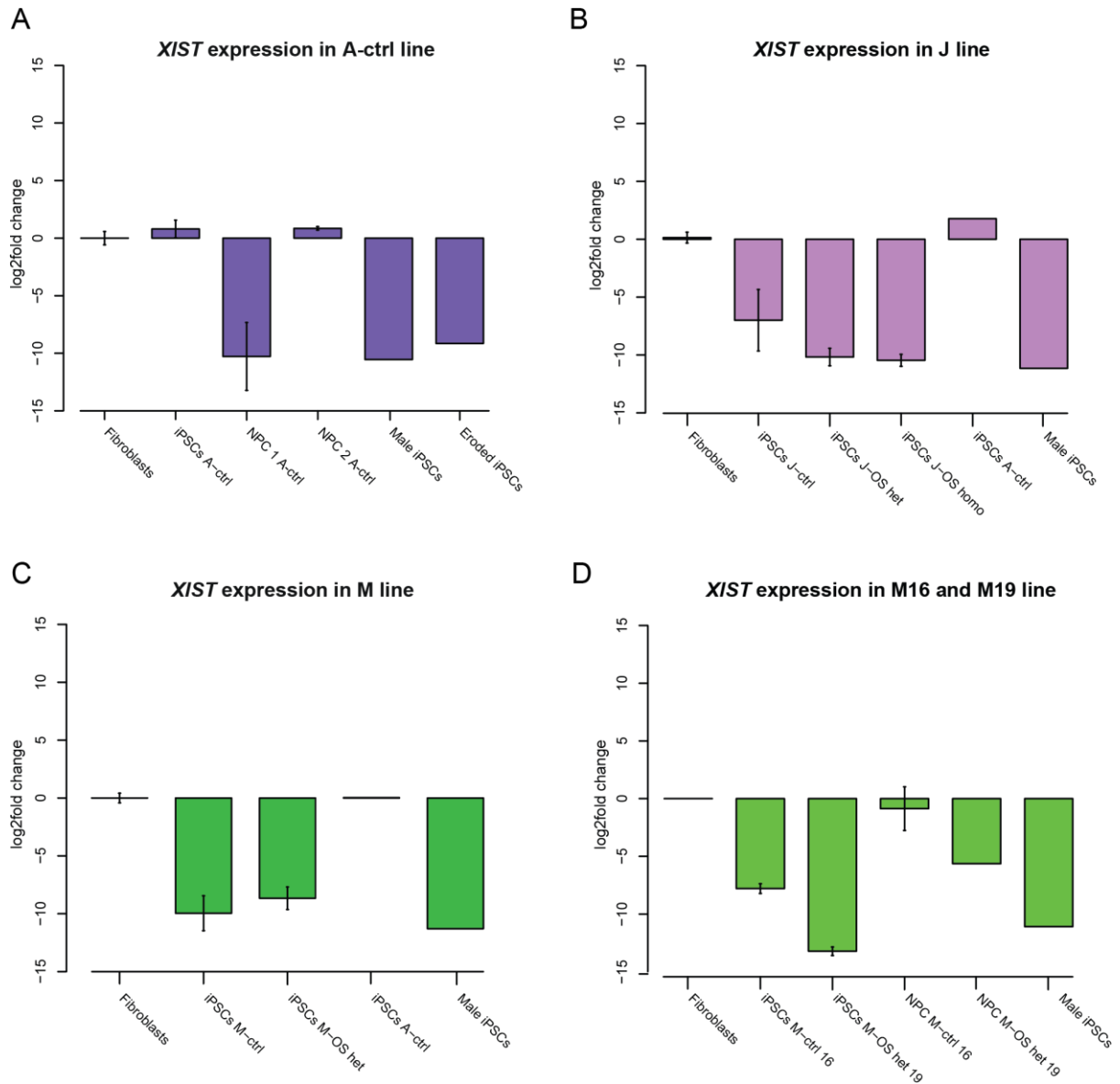


Figure 31. *XIST* expression monitored via RT-qPCR confirmed erosion and non-erosion status. RT-qPCR analysis for *XIST* gene in the cell lines used in the study. Fibroblasts, male iPSCs, eroded and non-eroded iPSCs were used as controls for *XIST* presence and absence. A) Fibroblasts from which iPSCs are derived were used in the experiment. n=3 for fibroblasts, iPSCs A-ctrl, NPCs 1 A-ctrl, and NPCs 2 A-ctrl; n=1 for male iPSCs and eroded iPSCs. B) Fibroblasts used for J-line reprogramming were used as controls. n=3 for fibroblasts; n=4 for iPSCs J-ctrl; n=2 for J-OS het and J-OS homo; n=1 for iPSCs A-ctrl and male iPSCs. C) J-ctrl fibroblasts were used as controls. n=3 for fibroblasts; n=5 for iPSCs M-ctrl; n=4 for iPSCs M-OS het; n=1 for iPSCs A-ctrl and male iPSCs. D) J-ctrl fibroblasts were used as control. n=3 for M-ctrl 16 iPSCs and NPCs, and for M-OS het 19; n=1 for fibroblasts, NPCs M-OS het 19, and male iPSCs.

Erosion can influence gene expression from the inactive X-chromosome, and it is associated with the passaging of cells⁴⁶⁹⁻⁴⁷³. During extensive culturing of iPSCs, *XIST* expression can be lost, leading to cell erosion. To understand the effect of erosion on the reactivation of gene expression from the inactive X-chromosome, we monitored the allelic expression of *MIDI*, a reactivated candidate gene, in eroded (M-ctrl, M-OS het, J-OS het) and non-eroded (M-ctrl 16) iPSCs and NPCs. iPSCs were analysed between passage 20 and 25, when erosion is not yet complete and some cells retain *XIST* expression,

and passage 40, where cells have *XIST* expression levels similar to male iPSCs, indicating complete loss of *XIST* expression and erosion. We first confirmed the trend of *XIST* expression in these samples via qPCR (Figure 32.A). Allele-specific RT-PCR for *MIDI* transcripts was performed on the same samples (Figure 32.C). The gel showed strict monoallelic *MIDI* expression throughout the passaging of the M-ctrl and M-OS het iPSC clones, indicating that *MIDI* is not subject to the artificial biallelic expression induced by the erosion mechanism. The same was observed for the J-OS het iPSC samples (Figure 32.B, C): J-OS het iPSC samples showed monoallelic *MIDI* expression despite erosion status and the complete absence of *XIST* expression, confirming in an additional cell line that erosion is not responsible for the reactivation we observed. To further understand whether *MIDI* reactivation is associated with the erosion process, allele-specific RT-PCR was performed on M-ctrl 16 iPSCs, which have an intermediate degree of erosion, and *XIST*-positive NPCs (Figure 31.D). The gel results showed that *MIDI* is expressed from both the active and inactive X-chromosomes in NPCs but not in iPSCs, confirming that erosion does not influence *MIDI* reactivation (Figure 32.D).

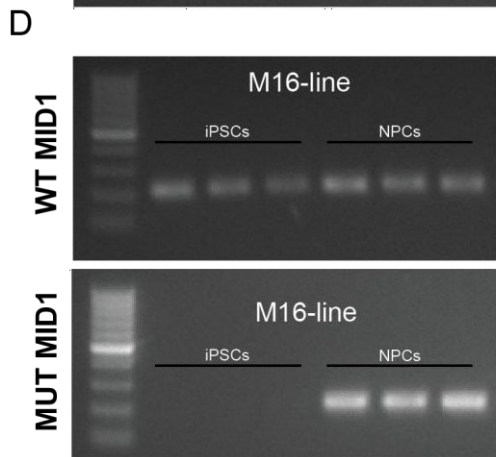
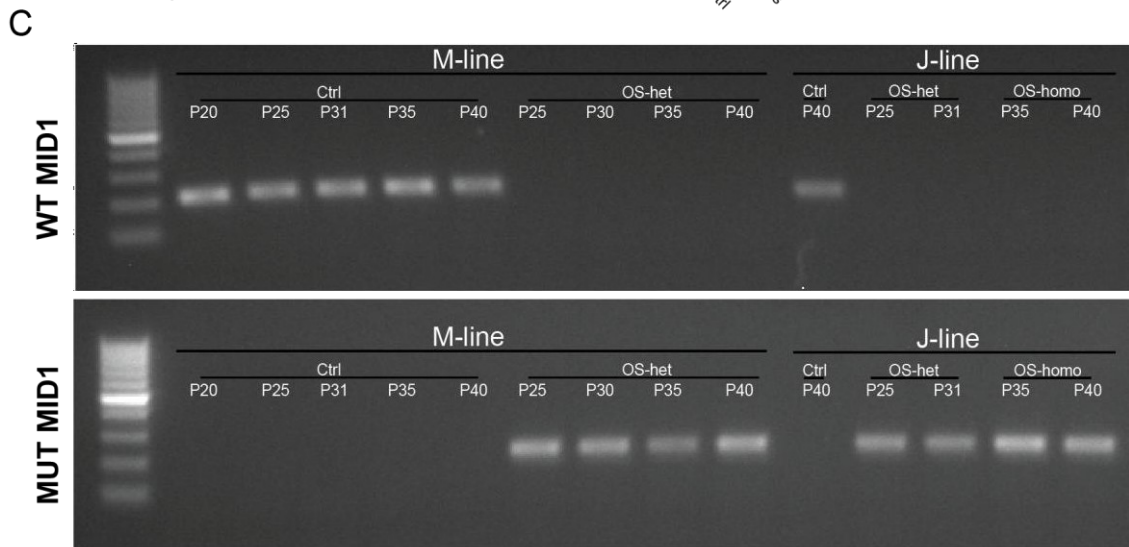
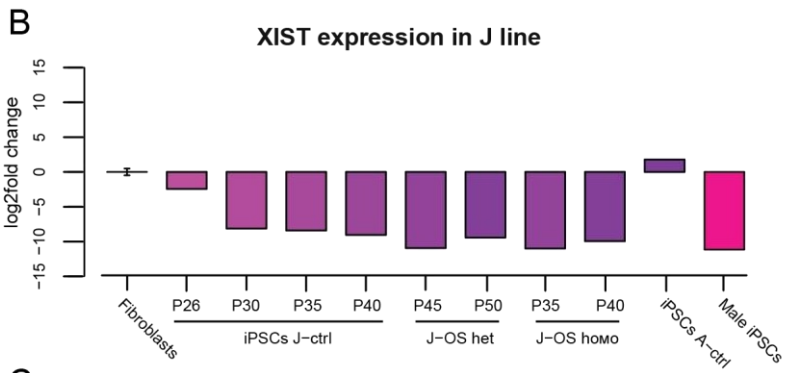
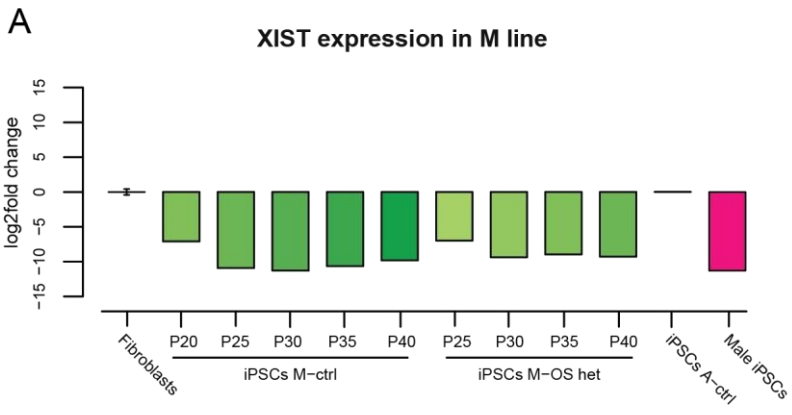


Figure 32. *XIST* expression during iPSC culturing and allele-specific *MIDI* RT-PCR in eroded iPSCs and non-eroded M-ctrl 16. RT-qPCR *XIST* analysis during passaging of M-line iPSCs (A) and J-line iPSCs (B). n=3 for fibroblasts; n=1 for other samples. C) Allele-specific RT-PCR for *MIDI* transcripts in iPSC samples collected at different stages of culturing. Specific passage number is shown in the figure (for example, P25 is passage 25). Same samples were tested for *XIST*-expression (Figure 32.A, B). Gel results showed that erosion associated with iPSC samples does not lead to biallelic *MIDI* expression detected in NPCs. B) Agarose gel results of allele-specific RT-PCR confirmed the *MIDI* biallelic expression in *XIST*-positive M-ctrl 16 clone. Same iPSCs and NPCs samples were first tested for *XIST*-expression via qPCR (Figure 31.D).

4.3.3 Single-nuclei validation of dynamic expression from the inactive X-chromosome by using RNA-FISH

The next step of the project aimed at validating dynamic expression from the inactive X-chromosome at single-cell resolution. All the methods employed so far are based on average expression analyses and do not consider single-cell variability or the complexity of the in vitro system used. To experimentally validate the dynamic expression from the inactive X-chromosome, we employed RNA-FISH technology, which allowed us to investigate the expression of gene candidates at single-nucleus and single-allele resolution. Non-eroded A-ctrl iPSCs and NPCs were used for RNA-FISH experiments, allowing us to accurately follow gene expression from the inactive X-chromosome territory in the nucleus, identified by *XIST* RNA staining (Figure 33.A). Allelic gene expression of X-linked candidates was followed by using RNA-FISH probes for pre-mRNA molecules, the unspliced RNA transcribed from the chromosomes and still associated with the site of transcription. By detecting the pre-mRNA molecule, we were able to associate gene transcription with the specific X-chromosome from which the gene is transcribed, either the active X-chromosome (without *XIST* cloud) or the inactive X-chromosome (with *XIST* cloud). Probes for pre-mRNA targets were designed using the bioinformatic online tool from ACD Biotechnie.

To confirm the functionality of the designed probes, control samples were treated with RNase A before probe hybridization. This confirmed that in the RNase A-treated samples, the probes were unable to hybridize with degraded RNA, thus preventing a non-specific fluorescent signal (Figure 33.C). An additional negative control was performed by using dilution buffer during the probe incubation step, for assessing the background fluorescence of the method (Figure 33.C). Both negative controls were performed for each RNA-FISH experiment. The RNA-FISH technology used for the experiment was RNAScope, a semi-quantitative RNA-FISH method based on amplification steps to highlight single RNA molecules. The dimension of the dots represented the quality of the RNA molecules.

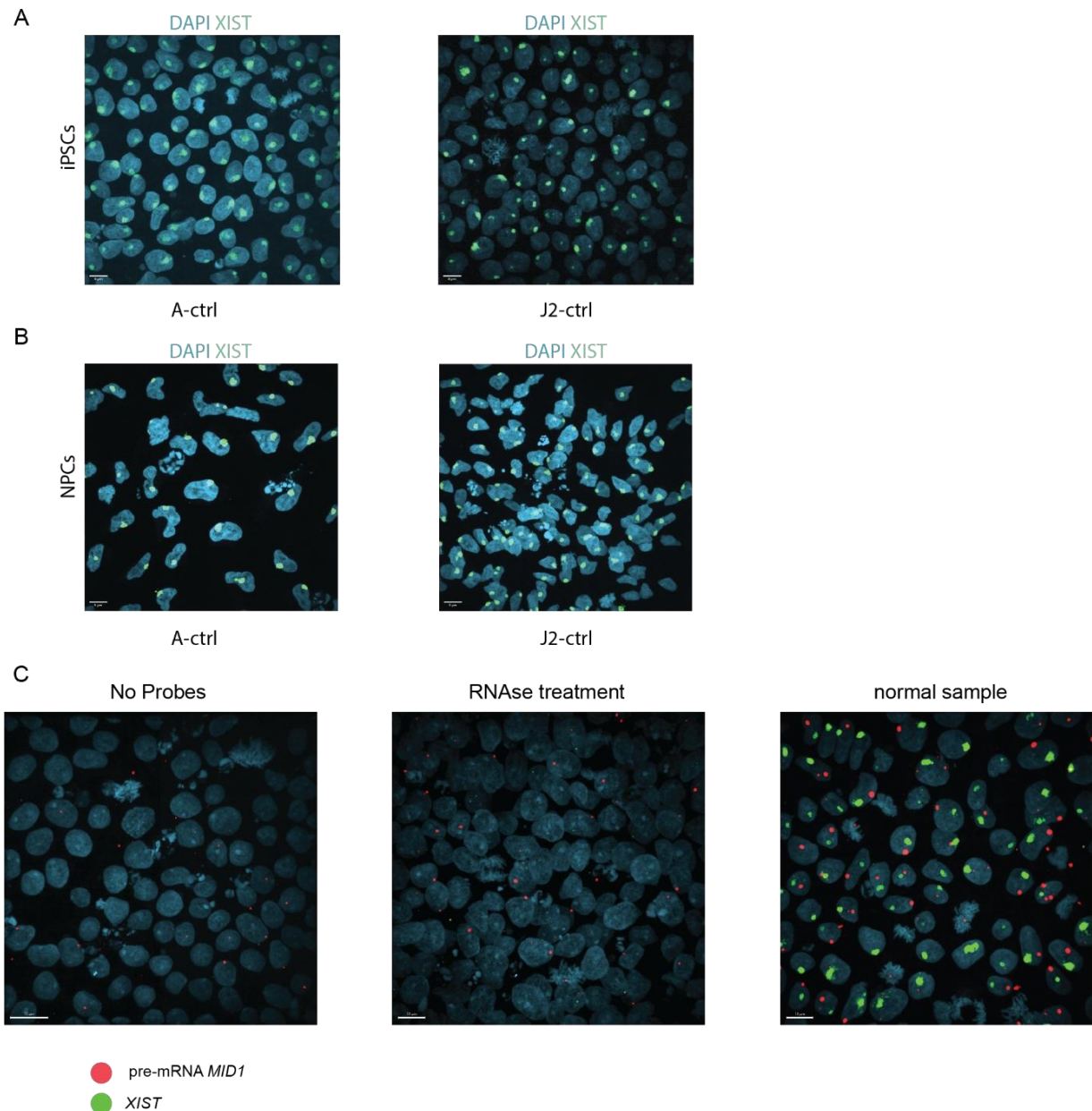


Figure 33. *XIST*-positive clones used for RNA-FISH experiments and negative controls compared to normal sample. *XIST* RNA-FISH representative pictures for iPSCs (A) and NPCs (B) showing the majority of the nuclei with a clear *XIST* signal coating the inactive X-chromosome. Scale bar = 9 μ m. C) Representative images of negative controls performed during the RNA-FISH protocols compared to normal treated sample. Note that the signal in the first two pictures derives from aspecific background. Scale bar = 15 μ m.

For the RNA-FISH experiment, we used probes targeting inactive, full-escape, and reactivated genes. Inactive genes and full-escape genes served as negative and positive controls, respectively, to confirm that the RNA-FISH technology is able to detect the presence of nuclei with a monoallelic or biallelic signals coming from the X-chromosomes. Furthermore, the monoallelic/biallelic signal ratio for the inactive and full-escape genes was expected to be consistent between iPSCs and NPCs.

Inactive genes were selected based on allelic predictions from the RNA-seq data and their expression values in iPSCs and NPCs. *SPIN3* was chosen as candidates for RNA-FISH experiments. The full-

escape gene *NLGN4X* was chosen due to its high expression levels in both iPSCs and NPCs, and previous studies that has identified *NLGN4X* as an escapee gene⁴⁶⁴. *MIDI* and *GPM6B* were selected as reactivated candidates. After designing probes for the pre-mRNA of those genes, RNA-FISH experiments were performed, and images were acquired using confocal microscopy (see section 4.2.8.2). The images were then analysed using the Imaris program, allowing for 3D visualization of the confocal data (see section 4.2.8.3).

XIST-positive nuclei and pre-mRNA positive nuclei were considered for analysis. The percentage of *XIST*-positive cells was calculated based on the number of *XIST*-positive nuclei, which displayed a clear inactive X-chromosome territory, divided by the total number of nuclei. The percentage of *XIST*-positive nuclei confirmed that in the A-ctrl, erosion is occurring in a small subset of cells, with the majority of iPSC samples showing more than 75% *XIST*-positive nuclei for the A-ctrl cell line (Figure 34). The percentage of *XIST*-positive cells slightly increased in NPCs, indicating a selection of *XIST*-positive cells during iPSC differentiation into NPCs, as reported in the literature⁴⁷⁴.

Pre-mRNA positive cells varied depending on the specific probe used, indicating that the expression level of the target genes influences the number of positive cells detected via RNA-FISH. Furthermore, detecting pre-mRNA species reduced the number of positive cells, as pre-mRNA is a small fraction of the total mRNA. Highly expressed genes, such as *NLGN4X* or *MIDI*, showed percentages of positive nuclei around 60% (Figure 34.A, D), while the percentage of pre-mRNA positive nuclei for low expressed *SPIN3* gene dropped below 50% (Figure 34.C).

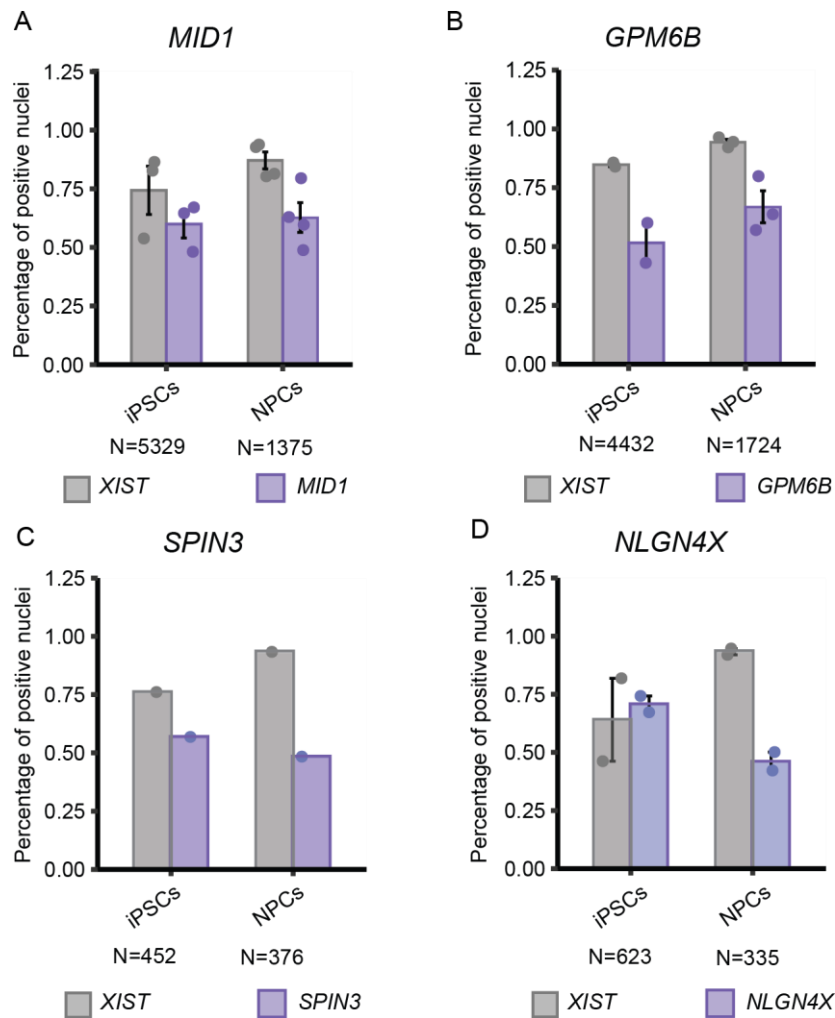


Figure 34. Percentage of *XIST*-positive and pre-mRNA-positive cells in A-ctrl RNA-FISH experiments. Barplots representing *XIST*-positive and pre-mRNA positive cells in A-ctrl RNA-FISH results. Dots represent the biological replicates for each experiment. N represents the number of total iPSCs and NPCS nuclei analysed for each RNA-FISH experiment.

Figure 35 presents representative images and quantification analyses for negative and positive probes. Representative pictures, as well as quantification analyses, were performed using the Imaris program. For quantification, only *XIST*-positive nuclei with a clear inactive X-chromosome territory were considered. Nuclei with more than two pre-mRNA signals, or with two pre-mRNA signals where one of the signals is not colocalizing with *XIST* territory, were excluded from the analysis.

RNA-FISH quantification of negative control probe signal showed no significant differences between the ratio of monoallelic to biallelic populations in iPSCs and NPCs. Furthermore, the percentage of biallelic populations for *SPIN3* decreased from iPSCs to NPCs, although this decrease was not statistically significant.

Quantification of *NLGN4X* signal distribution showed a high number of nuclei with a biallelic signal in both iPSCs and NPCs, indicating no significant difference between these cell types. This result highlights that not all nuclei exhibited biallelic expression even when full-escape gene were analysed in both iPSCs and NPCs.

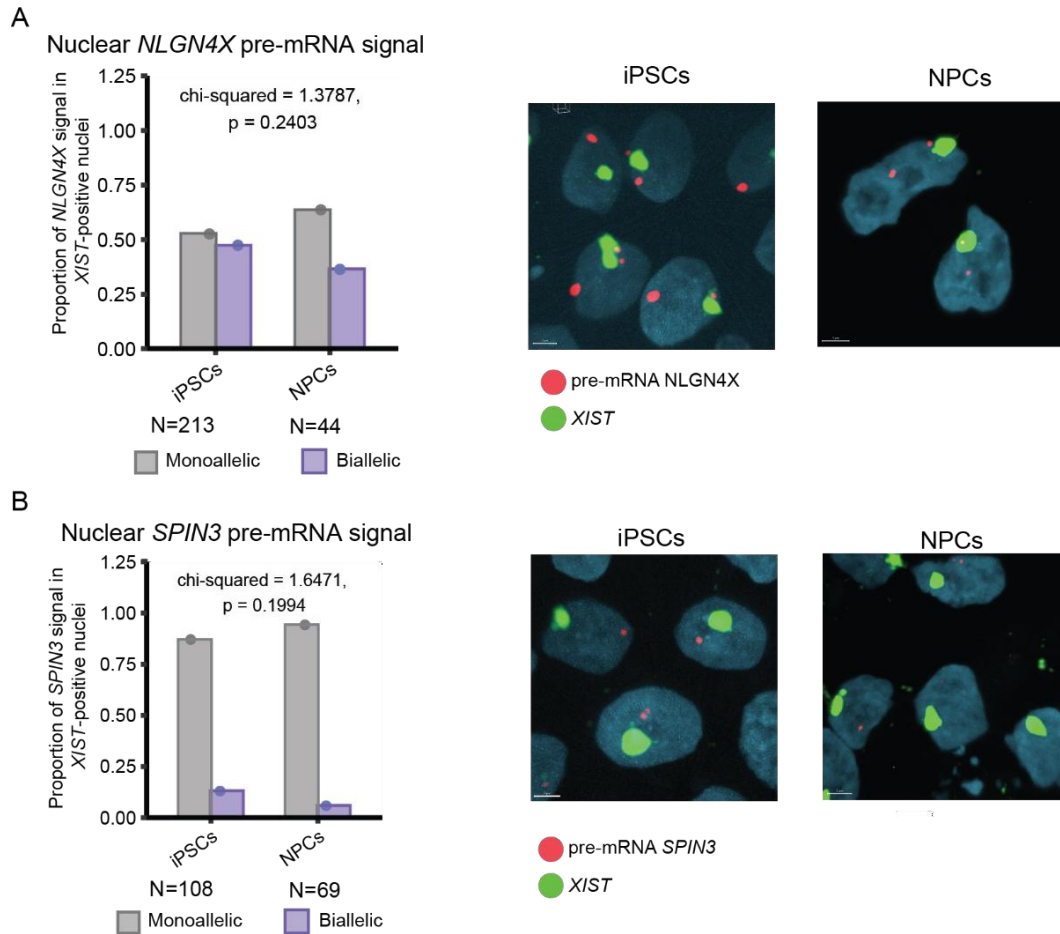


Figure 35. Full-escapee and inactive genes showed similar monoallelic-biallelic nuclei distribution between iPSCs and NPCs. RNA-FISH quantification results and representative images for full-escapee *NLGN4X* gene (A), and inactive genes *SPIN3* (B) in A-ctrl iPSCs and NPCs. Dots represent the number of biological replicates for each experiment, while N represents the number of total *XIST*-positive nuclei with a clear monoallelic or biallelic pre-mRNA signal considered for the analysis. Samples were statistically compared by using the chi-squared test, and p-values were reported above the barplots. Scale bar = 3 μm for A, Scale bar = 2 μm for B.

To further investigate the reactivation mechanisms at single-nuclei resolution, we designed probes for the pre-mRNA of the *MIDI1* and *GPM6B* genes and performed RNA-FISH experiments using *XIST*-positive A-ctrl iPSCs and NPCs. Quantification showed a significant increase in the number of nuclei with a biallelic signal in NPCs compared to iPSCs for both *MIDI1* and *GPM6B* (Figure 36).

By comparing the nuclear distribution of the pre-mRNA signal between iPSCs and NPCs for the inactive, full-escapee, and reactivated genes tested, significant changes in the monoallelic/biallelic ratio could only be observed for the reactivated genes tested. This highlights that the RNA-FISH technology can capture dynamic gene expression from the inactive X-chromosome. The results suggest that reactivation of gene expression from the inactive X-chromosome could be confirmed at the single-nuclei level during neural differentiation of pluripotent cells.

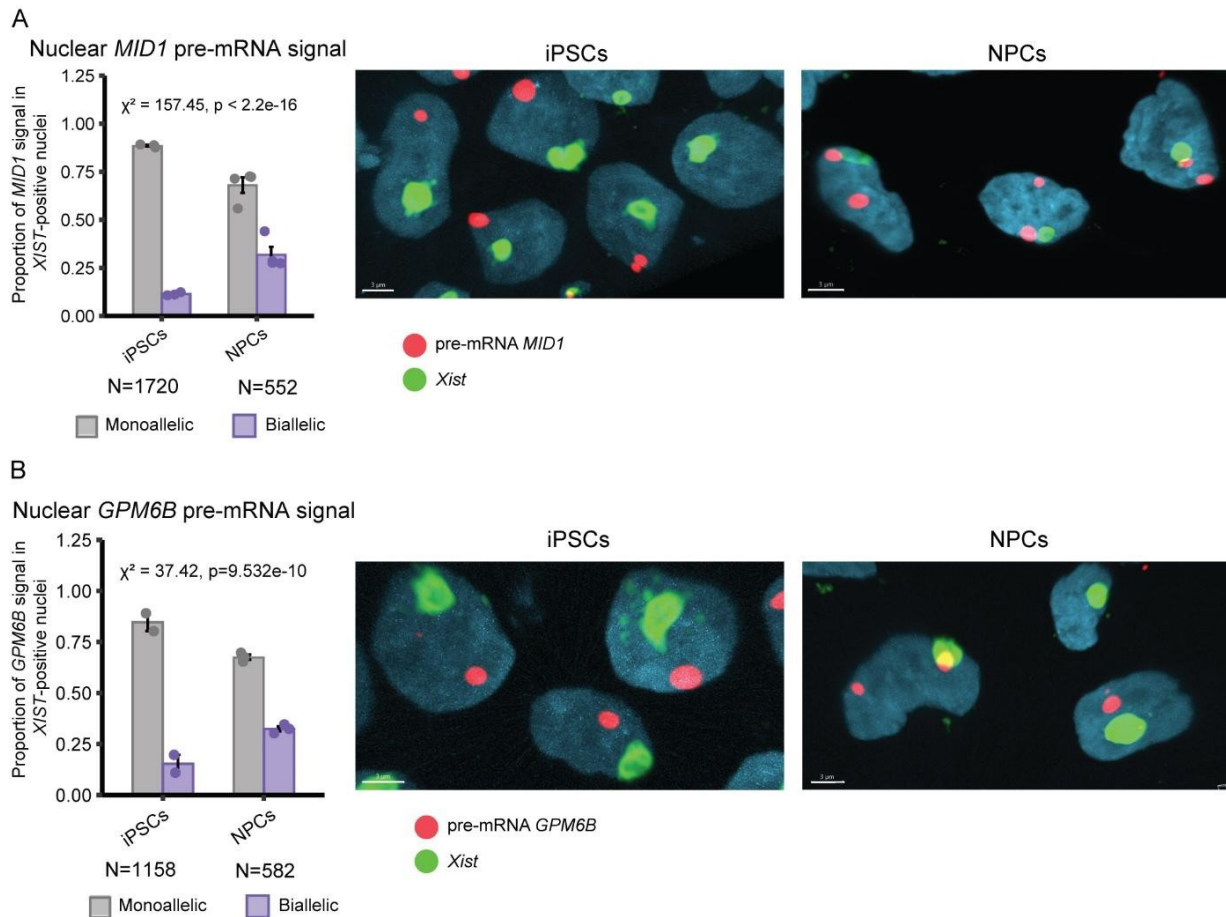


Figure 36. Reactivated genes *MID1* and *GPM6B* showed a significant increase in the biallelic nuclei population in NPCs compared to iPSCs. RNA-FISH representative images and quantification analysis for the reactivated genes *MID1* (A) and *GPM6B* (B) in A-ctrl iPSCs and NPCs. Dots in the barplots represent the biological replicates of the experiments, while N represents the number of total *XIST*-positive nuclei with a clear monoallelic or biallelic pre-mRNA signal considered for the quantification analyses. Samples were statistically compared by using the chi-squared test, and p-values were reported above the barplots. Scale bar = 3 μm .

Since the J-ctrl iPSC clone used for the RNA-seq experiment showed extensive erosion of the inactive X-chromosome, additional J-ctrl iPSC clones were screened and a high-*XIST* expressing J-ctrl iPSC clone was found. Successful confirmation of the non-eroded J2-ctrl clone was achieved by *XIST* RNA-FISH, and NPCs were subsequently generated (Figure 37). RNA-FISH experiments were performed using non-eroded J2-ctrl iPSCs and NPCs. Only the reactivated *MID1* and *GPM6B* genes were investigated with RNA-FISH. The percentage of positive cells for the pre-mRNA probes was similar to those calculated for the A-line (Figure 38). Interestingly, the percentage of *XIST*-positive cells was lower than in the A-line, suggesting that even though early passage cells were used for the experiments (passages fewer than 15), erosion occurred at single-nuclei level in J-line. This finding aligns with qPCR and H3K27me3 results (Figure 30 and Figure 32.B).

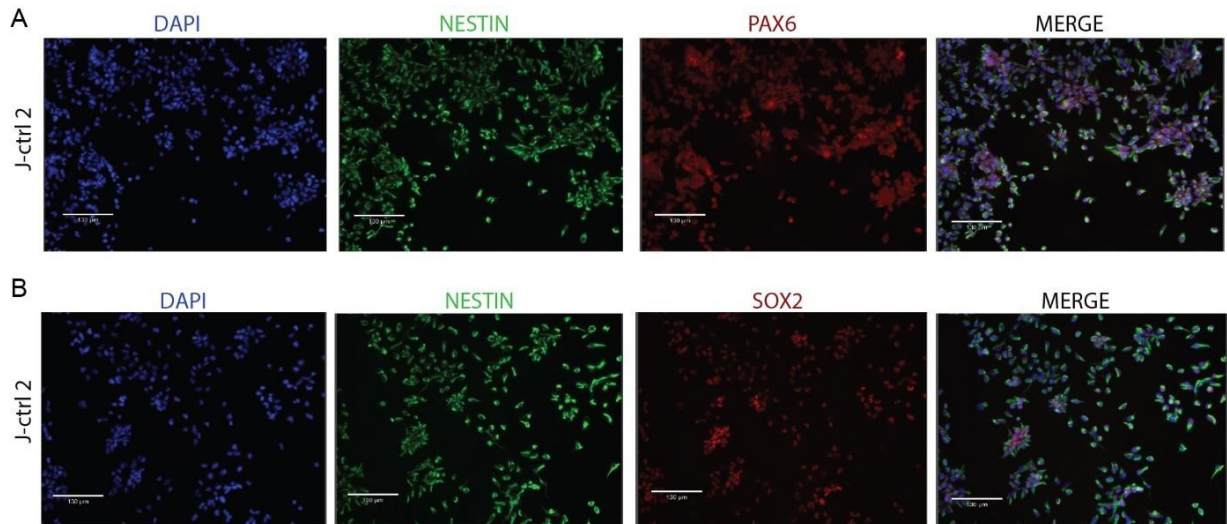


Figure 37. Confirmation of J2-ctrl NPCs differentiation. *XIST*-positive J2-ctrl iPSCs were differentiated into NPCs by following the established workflow. Immunostaining analysis confirmed the presence of the protein markers PAX6, SOX2, and NESTIN. Scale bar = 130 µm.

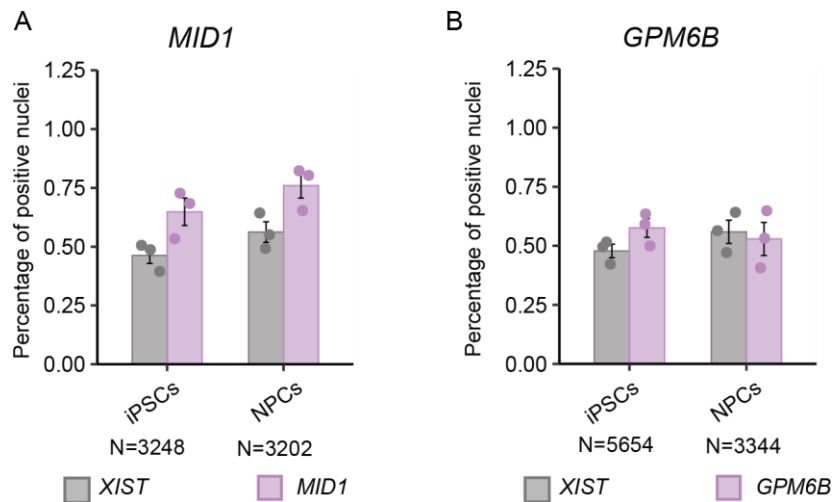
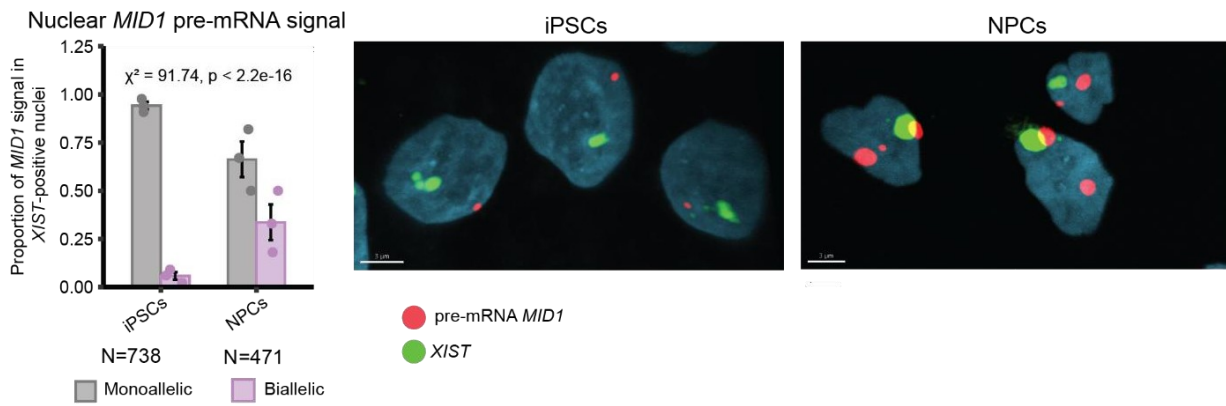


Figure 38. Percentage of *XIST*-positive and pre-mRNA-positive cells in A-ctrl RNA-FISH experiments. Barplots representing *XIST*-positive and pre-mRNA positive cells in J-ctrl 2 RNA-FISH experiments. Dots represent the biological replicates for each experiment. N represents the number of total iPSCs and NPC nuclei for each RNA-FISH experiment.

Quantification of monoallelic and biallelic nuclei in J2-ctrl with *MID1* signal showed a significant increase of biallelically expressing nuclei in NPCs compared to iPSCs, with fewer than 6% of nuclei showing a biallelic signal in iPSCs, and more than 33% in NPCs (Figure 39.A). For *GPM6B*, the percentage of biallelic nuclei did not change between iPSCs and NPCs, with fewer than 20% of the biallelic population in both iPSCs and NPCs (Figure 39.B). Interestingly, NPCs with a higher passage showed a higher percentage of biallelic nuclei for both *MID1* and *GPM6B*, suggesting that the reactivation process increases with the passage of the NPCs (Figure 39). A conclusion of the results

A



B

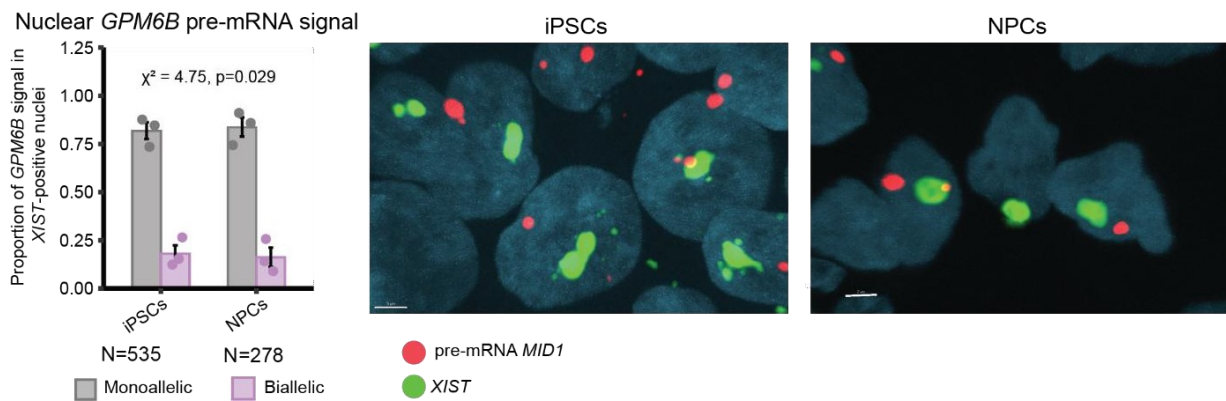


Figure 39. Reactivated genes *MID1* and *GPM6B* monoallelic to biallelic proportion in J-ctrl 2 iPSCs and NPCs. RNA-FISH representative images and quantification analysis for the reactivated genes *MID1* (A) and *GPM6B* (B) in J-ctrl 2 iPSCs and NPCs. Dots in the barplots represent the biological replicates of the experiments, while N represents the number of total *XIST*-positive nuclei with a clear monoallelic or biallelic pre-mRNA signal considered for the quantification analyses. Samples were statistically compared by using the chi-squared test, and p-values were reported above the barplots. Scale bar = 3 μ m, Scale bar = 2 μ m for B) NPCs.

5 Discussion

Dosage compensation of the sex chromosomes ensures the balance of X-linked gene expression between males (XY) and females (XX) by silencing each extra copy of the X-chromosome in female eutherian mammals. The silencing of the additional X-chromosome in females is a conserved dosage-compensation mechanism across mammals. However, only recently have the details of this process begun to be fully understood in different mammalian species. The mechanisms used by various species to achieve X-linked dosage compensation differ, with lncRNAs playing a key role in the process^{475–478}. Despite these differences, X-chromosome inactivation (XCI) is a stable epigenetic event, where one random X-chromosome is transcriptionally silenced and this silencing is inherited to the daughter cells. In humans, the additional X-chromosome is thought to at least partially account for the sex differences observed in both physiological and pathological conditions^{479–483}. This is because approximately 20% of X-linked genes continue to be expressed from the inactive X-chromosome^{301,465,484}. These genes are known as escape genes and are categorized into two types: genes that always escape from the inactive X-chromosome called constitutive escapees, and genes that are expressed from the inactive X-chromosome only in a subset of individuals, tissues, or in a specific temporal stage of life. Those genes are called facultative escapees^{484,485}. The transcriptional activity of the inactive X-chromosome may help explain the sex bias observed in disease susceptibility in humans, as many genes located on the X-chromosome are involved in immune and nervous system functions⁴⁴². Facultative escapees, in particular, could play a key role in disease-related sex bias due to their tissue- or temporal-specific expression from the inactive X-chromosome. This makes them part of a dynamic system that can bypass dosage compensation for specific X-linked genes in particular tissues or developmental stages. However, while studies identified brain-specific facultative escapees in adult mice and during *in vitro* neural differentiation^{448,463,464}, there are no studies examining the dynamics of facultative escapees during embryonic differentiation or brain organogenesis in humans.

During my PhD thesis, I leveraged the clonal status of iPSCs with skewed XCI together with *in vitro* 2D neuronal development to establish an *in vitro* system for the detection of facultative and constitutive escape genes during neuronal differentiation. By analysing transcriptomic data at allele-specific resolution, we identified novel candidates for constitutive escape genes, as well as new categories of facultative escapees that change their expression from the inactive X-chromosome depending on the cell type and differentiation status. We identified three distinct categories of escape genes: full escape genes, which likely correspond to constitutive escapees by consistently evading transcriptional silencing of the Xi in both iPSCs and throughout neural differentiation. Furthermore, we identified two novel categories of facultative escape genes not previously described in the literature. These included reactivating genes, which escape from the Xi only in neuronal cells but not in iPSCs; and late-silenced genes, which escaped in iPSCs but become silenced in NPCs and neurons. Furthermore, using RNA-FISH technology at single-cell resolution, I confirmed the dynamic expression of the reactivated candidate genes *MIDI* and *GPM6B* during the differentiation of iPSCs into NPCs.

5.1 *In vitro* differentiation of human pluripotent stem cells can faithfully recapitulate the human neurodevelopmental process

Induced Pluripotent Stem Cells (iPSCs) were first introduced by Yamanaka as a type of Pluripotent Stem Cells (PSCs) derived from somatic cells through the induction of OSKM factors expression⁴⁸⁶. Since their discovery, iPSC technology has enabled basic researchers to study cell types and developmental processes that cannot be obtained from primary material, such as human neural development and neurons. Additionally, iPSCs are widely used by clinical researchers to model genetic disorders by generating crucial disease-relevant cell types from patient-derived primary cells.

5.1.1 Reprogramming of somatic cells into iPSCs

In my PhD study, I utilized induced pluripotent stem cells (iPSCs) as an in vitro model of pluripotency to replicate human neural development and explore changes in gene expression from the inactive X-chromosome during neuronal differentiation. Human fibroblasts, sourced from both patients and healthy controls, were reprogrammed using the commercially available kit based on the Sendai virus vector (CytoTune™-iPS 2.0 Sendai Reprogramming Kit). Prior to reprogramming, fibroblasts were tested and confirmed negative for human viral pathogens (HIV, HBV, and HCV). The reprogramming procedure, along with the workflow for cell characterisation, had been previously established (Thesis, Dr. Stephan Käseberg). This method involved the ectopic expression of the genes *KLF4*, *Oct3/4*, *Sox2*, and *c-Myc* in somatic cells, leading to major remodelling of the epigenome, transcriptome, proteome, metabolome, and overall cell biology, resulting in the generation of pluripotent stem cells from terminally differentiated cells⁴⁸⁷. Following reprogramming, the iPSC lines were rigorously characterized for morphology and pluripotent state, and karyotype analysis was conducted to rule out any chromosomal aneuploidies or rearrangements caused by the reprogramming procedure⁴⁸⁸. The iPSCs were compared with their fibroblast cells of origin and with other established iPSC lines in the laboratory. Additionally, before proceeding with further experiments, the removal of the Sendai Virus vector was verified to ensure that the cells' biology was not influenced by the presence of viral vectors, thereby minimizing associated tumorigenic risks for the iPSC user and to avoid that SeV vectors influenced experimental results⁴⁸⁹. The long-term presence of SeV vectors could impair the differentiation potential of the iPSCs by inducing ectopic long-term expression of OKSM factors⁴⁹⁰. Despite their pluripotency and self-renewal capabilities, extensive culturing of iPSCs can lead to clonal selection of mutations that might alter their biology, such as enhanced proliferation due to mutations in oncogenes⁴⁹¹. To mitigate this risk, the iPSCs used in the study were maintained at a passage number no greater than 45, thereby reducing the likelihood of chromosomal aberrations.

In the first paragraph of the Results section, I described the characterization of the 1262/16 iPSC J-line. The reprogramming of this line was completed prior to the start of my PhD, while I conducted the subsequent characterization using immunostaining and qPCR techniques. Additionally, karyotype analysis was performed on the iPSC clones used in the study, confirming the absence of chromosomal aberrations. The absence of the Sendai virus vectors used during reprogramming was confirmed by RT-PCR by following the company's protocol. This ruled out the possibility that the episomal vectors used in the reprogramming process could alter the iPSC cell biology^{490,492}. The resulting human iPSCs exhibited characteristics similar to embryonic stem cells, particularly resembling epiblast cells at the post-implantation stage, which is known as the primed state^{487,493-495}. In this state, the FGF/Activin signalling pathway plays a crucial role in maintaining stem cell identity⁴⁹⁶⁻⁴⁹⁸. Cells in this embryonic stage display a flat morphology with round colonies characterized by sharp edges, matching the morphology of the iPSCs⁴⁸⁷.

Representative images from the immunostaining experiments showed that all nuclei were positive for the transcription factor NANOG, a protein expressed in the inner cell mass of the blastocyst that is crucial for maintaining pluripotency and stem cell renewal by suppressing fate-determination transcription factors^{499,500}. Additionally, the surface glycoprotein Tra 1-60, commonly expressed in hESCs and hiPSCs⁴⁹⁹, was also detected. A comparison of iPSC markers between the J-line and the A-line, another iPSC line previously characterized in the laboratory, revealed no significant differences in the expression of *OCT4*, *KLF4*, and *NANOG*. This confirmed the similarity between the J-line and the previously characterized A-line iPSCs, indicating the high quality of the J-line iPSCs and supporting their suitability for further experiments.

In conclusion, a robust method for generating iPSCs from human dermal fibroblasts was established in the lab prior to the start of my thesis (Thesis, Dr. Stephan Käseberg). The characterization workflow for these cells was meticulously established, incorporating a combination of different techniques and approaches. This resulted in the generation of high-quality hiPSCs that faithfully recapitulate a specific human embryonic stage, making them a suitable model for human embryonic development in my research. However, despite the gold-standard methods are based on previous publications using iPSCs, the field of hPSC research still lacks the establishment of official guidelines, leading to variability in the characterization of the pluripotent stem cells. Moreover, new methods are continually developed to better assess the pluripotency status of reprogrammed cells⁵⁰¹⁻⁵⁰⁵. In that regard, it will be essential to update the characterization workflow in the future by incorporating recent advancements in the iPSC characterization, such as flow cytometry analysis to assess the expression of pluripotency-associated markers, and the use of a trilineage differentiation kit to evaluate the cells' capability to differentiate into the three embryonic germ layers⁴⁸⁸.

5.1.2 iPSC differentiation into NPCs

Once the pluripotent identity of the iPSCs was confirmed, the cells were differentiated into neural progenitor and stem cells (NPCs) using a commercially available kit from Thermo Fisher Scientific. The NPC differentiation of iPSCs with this kit was established in the laboratory prior to the start of my PhD study, and it is widely used in the literature to obtain neural stem and progenitor cells for studying the early stages of neuronal differentiation^{506,507}. The differentiation protocol relies on dual SMAD inhibition, which directs pluripotent stem cells to develop into neuroectoderm cells and, subsequently, into neural progenitor cells^{508,509}. The resulting NPCs were characterized based on their morphology and the expression of neuronal progenitor cell-specific markers, as well as the downregulation of iPSC-specific markers. After successful characterization, these NPCs were established as stable proliferating cell lines for further experiments.

The expression of three specific markers was assessed at both the gene expression and protein levels. One of these markers is NESTIN (Neuroepithelial Stem Cell Protein), an intermediate filament protein expressed during the development of the central nervous system (CNS). NESTIN is present in both adult and embryonic CNS stem/progenitor cells and is downregulated as progenitor cells differentiate into neurons or glial cells⁵¹⁰⁻⁵¹⁴. Another marker is Pax6, a transcription factor from the paired box (Pax) family, which plays a critical role in eye development and brain organogenesis. Pax6 is involved in the proliferation and differentiation of neural stem and progenitor cells in various organisms, including flies, mice, and humans⁵¹⁵. The third marker analysed in the thesis is SOX2, a transcription factor from the SOX gene family, primarily associated with stem cell biology. SOX2 is expressed in the blastocyst stage in both the Inner Cell Mass and the Trophectoderm cells. During gastrulation, SOX2 is linked with differentiating ectoderm, mesoderm, and endoderm⁵¹⁶. *In vitro*, SOX2 is detected during the transition from embryonic stem cells to neuroectoderm. A similar role is observed *in vivo*, where, in the absence of other competing factors, SOX2 can induce stem cell differentiation into neural plate cells rather than paraxial mesoderm⁵¹⁷. During CNS development, SOX2, along with SOX1 and SOX3, regulates the proliferation of neural progenitor cells by repressing differentiation stimuli of the progenitor cells⁵¹⁸. All three markers were confirmed to be expressed at high levels in the differentiated NPCs, both at the gene expression and protein levels. The modest downregulation of *KLF4* in NPCs compared to iPSCs reflects the stem cell status of the neuronal progenitors, consistent with extensive literature reporting NPCs expressing the SKM factors⁵¹⁹.

As for the iPSCs, the characterization of the NPC cell identity was performed using a combination of different methods to ensure the proper differentiation into neural progenitor cells. Although using a commercially available kit, the outcome of the iPS differentiation into NPCs depends on several factors

including the quality of the iPSC cells and adaptation of the protocol made during the NPC differentiation depending on the state of the cells. This underscores the need for thorough assessment to ensure proper differentiation.

Similar techniques were employed for the characterization of both iPSCs and NPCs. Among them, the immunostaining for the NPC markers allowed for single cell visualization and revealed a small population of cells that were not positive for NPC specific markers, indicating the presence of falsely differentiated cells and that the differentiation protocol is not 100% efficient. This highlights the importance of immunostaining technique as a key method for the proper NPC cell identity assessment, as adopted in other studies⁵⁰⁶. Furthermore, the ability of the NPCs to differentiate into neurons further confirmed the proper identity of the neural progenitor cells.

5.1.3 NPC differentiation into neurons

To better model human neurodevelopment, NPCs were further differentiated into post-mitotic neurons by following an adapted protocol based on a Thermo Scientific procedure. The protocol was established before the start of my thesis (Thesis, Dr. Stephan Käseberg), and involved addition of retinoic acid (RA) to the media and subsequently removing the bFGF growth factor, which induces the differentiation of neuronal progenitor cells into neurons.

Neural stem cells can differentiate into various types of brain cells (neurons, astrocytes and oligodendrocytes) depending on the specific media composition used. The media formulation employed for neuronal differentiation in this study is similar to that used for maintaining neuronal stem cells, consisting of a serum-free formulation designed for culturing neurons⁵²⁰. The main difference between NPC and neuron media formulations is that, for neuronal differentiation, basic fibroblast growth factor (bFGF) is removed while retinoic acid is added. Neuronal stem cells are typically maintained in a proliferative state by the mitogen bFGF; however, removing bFGF from the media stimulates the cells to differentiate and exit the cell cycle. Additionally, the addition of retinoic acid enhances the ability of neural progenitor cells to exit the cell cycle due to RA's direct effect on cell cycle arrest, such as upregulating p27 and inhibiting the function of ZIC1 transcription factors^{521,522}. This approach aligns with the developmental concept that opposing signalling pathways regulate different stages of development⁵²³, and a similar strategy has been employed in established protocols for the neuronal differentiation of human pluripotent stem cells.

The success of neuronal differentiation was confirmed by analysing the expression of neuronal and NPC markers through qPCR and immunostaining, ensuring proper neuronal cell identity using different approaches. The expression of neuronal markers MAP2, TAU, and TUBB3 was specifically observed in the differentiated neurons. MAP2 is part of a class of proteins associated with microtubules and specifically expressed in neurons, particularly in dendritic structures⁵²⁴. *TUBB3* gene translates into the class III of β -Tubulin, which is specifically expressed in the brain and testis⁵²⁵. Finally, TAU protein is another microtubule-associated protein predominantly expressed in neurons and plays a role in proper cytoskeleton organization⁵²⁶. The expression of all these markers was confirmed at both the gene and protein levels, verifying that the neuron-like cells are indeed post-mitotic neurons.

One limitation of the differentiation paradigm used in this study was the unguided differentiation of NPCs into specific neuronal types and the lack of characterization of the neuronal subtypes present in the *in vitro* system. Preliminary analysis was conducted by examining the expression of neural markers from the RNA-seq dataset. The neuronal samples exhibited a mixed expression of both excitatory and inhibitory markers, indicating that the differentiation paradigm employed in the study does not selectively yield a specific neuronal population.

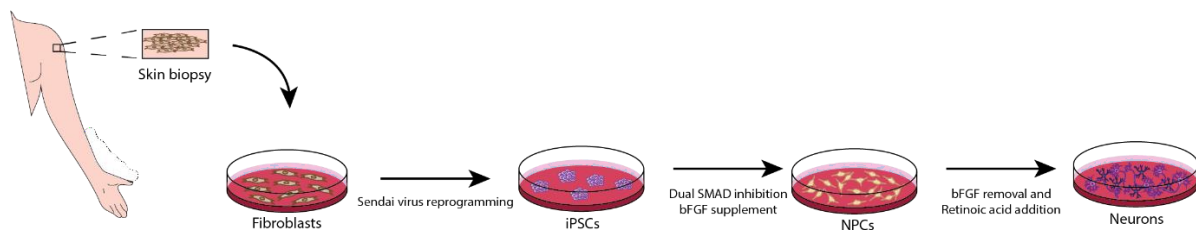


Figure 40. iPSC generation and 2D neural differentiation. Cartoon representing the fibroblasts isolation, iPSC reprogramming and the 2-step neural differentiation of iPSCs into NPCs and neurons.

5.2 Usage of the *in-vitro* system to study facultative escape from X-inactivation

The development of an *in vitro* system that faithfully recapitulates the stages of human neurodevelopment allowed the investigation of the facultative escape genes during neuronal differentiation. The clonal nature of the hiPSCs used in the study reduced the efforts in identifying the escape genes. However, challenges remain in translating these discoveries to their *in vivo* counterpart due to the inherent limitations of the *in vitro* system.

5.2.1 Skewed *in vitro* system can reliably identify facultative escape genes

Human iPSCs share the same X-chromosomal status of embryonic post-implantation epiblast cells, where X-chromosome inactivation (XCI) results in one randomly chosen X-chromosome being inactivated, leading to the presence of one active and one inactive X-chromosome⁵²⁷. Furthermore, studies in iPSCs have shown that most lines exhibit clonal XCI, meaning that the same X-chromosome remains consistently inactive while the other remains active⁵²⁸. This was also observed in the lines used in this study.

During the reprogramming process, colonies were picked and expanded, leading to the generation of clonal-derived iPSC lines. Pyrosequencing results for the expression of a heterozygous SNP in the X-linked *ZNF185* gene confirmed the difference between fibroblast samples, where random XCI creates a mosaic population with heterogeneous SNP expression, and the iPSC clones, where SNP expression originates from a single allele. However, the mosaic SNP expression in fibroblasts from the M and J-lines indicates significant skewing of the X-chromosome, with skewing levels ranging from 25% to 10%. In the literature, such values are considered indicative of fully skewed individuals, suggesting that the donors themselves have non-canonical XCI^{529,530}.

Further confirmation of the clonal status of the iPSCs was provided by allele-specific RT-PCR for the *MIDI* gene. A 4 bp deletion in one of the *MIDI* alleles in the M-line allowed for the specific amplification of either the wild type or the variant *MIDI* allele. These results confirmed the expected clonal nature of the iPSC clones tested.

The use of a clonal system as a starting point for allele-specific analysis is a well-established method in X-inactivation and escapee research, both in humans and mice^{301,448,463,531}. Leveraging skewed XCI enables the dissection of allele-specific expression signals from bulk RNA-seq data. *In vivo*, a clonal status can be achieved by using hybrid cells derived from crossing two mouse strains, where the active and inactive X-chromosomes are easily distinguished due to genomic variant differences. This approach was utilized by Berletch and group⁴⁴⁸. They performed allele-specific expression analysis in mouse

tissues by crossing *Mus musculus* and *Mus spretus* and using the F1 generation, characterized by a high SNP density in their genome (1 SNP per 50-100 bp). A mutation in the A-repeat of the *XIST* gene present on the *Mus musculus* X-chromosome led to a completely skewed XCI system in which, in the F1 generation, the *Mus musculus* X-chromosome failed to inactivate due to the *XIST* mutation, resulting in the complete inactivation of the *Mus spretus* X-chromosome. Using this approach, they were able to investigate the escape genes in various mouse tissues *in vivo* (brain, ovary, and spleen), identifying a subset of tissue- and cell type-specific escape genes.

Similarly, Giorgetti and colleagues⁴⁶³ performed chromatin profiling and gene expression analysis *in vitro* by using mouse embryonic stem cells (mESCs) and NPCs differentiated from them. mESCs were isolated from the highly polymorphic F1 generation derived from crossing the *Castaneous* with the *I29SI* mouse strain. In addition, after NPC differentiation of mESCs, clonal picking was performed, leading to complete skewness of XCI in NPCs, where the X-chromosome from either the *Castaneous* or the *I29SI* strains was widely inactivated in the specific NPC clones. This allowed them to unambiguously assign allele-specific signals to one or the other X-chromosome based on the strain-specific genetic variants.

Similar studies were performed in humans by Cotton and group⁵³¹. They identified skewed human samples based on the assumption that allele-specific expression levels of a set of silenced genes subject to the XCI process differ on average by 50%. They analysed allele-specific RNA-seq datasets derived from lymphoblastoid cell lines and fibroblasts to identify samples where the set of inactive genes had an allele-expression above the average 50% found across the genome. These cell lines were considered to have skewed XCI and were further used to assess the allele-expression levels of X-linked genes, identifying facultative escape genes that were cell line- and individual-specific, with variable expression levels ranging from minimal expression, sufficient to be considered as escape gene, to higher expression levels similar to autosomal genes.

Lastly, Tukiainer and colleagues used female expression data from a skewed individual³⁰¹. They identified a female donor where the allele-specific expression of X-linked genomic variants completely deviated from the average 50%, showing 100% - 0% ratio with no chromosomal abnormalities detected on the X-chromosomes. This unique dataset allowed them to investigate the XCI status of 186 X-linked genes, depicting an elaborate picture where half of the analysed genes showed variable expression from the inactive X-chromosome across 16 tissues, and even identifying a small set of genes showing biallelic expression in only one of the tissues analysed. The critical findings achieved in the studies mentioned above underscore the importance of using a fully skewed system for the detection of facultative escape genes.

Since fully skewed XCI individuals are rare⁵³², human *in vitro* systems become extremely valuable. *In vitro*, a clonal status can be achieved by picking and expanding a specific clone where one X-chromosome remains stably inactivated. This approach can be applied to both stem cells and differentiated cells, enabling the identification of cell type-specific facultative escape genes^{463,464}. However, no studies have yet tracked the expression of facultative escape genes during the differentiation stages of pluripotent cells into the neural lineage in humans. The fact that the iPSC clones used in this study were picked during the reprogramming process and were clonally derived provided a starting point for analysing X-linked gene expression along neuronal developmental trajectories.

5.2.2 Erosion of the inactive X-chromosome fails to explain the dynamic expression of facultative escape genes

A notable issue arising from the extensive *in vitro* culturing of iPSCs is the erosion of the inactive X-chromosome. During this process, the epigenetically silenced X-chromosome loses its *XIST* coating and

repressive epigenetic markers, leading to the ectopic expression of genes located on the inactive X-chromosome. The susceptibility of an iPSC line to erosion appears to be intrinsic to the line itself⁴⁷⁰, with a major risk factor being the number of passages the cells underwent *in vitro*⁴⁶⁷. To understand which factors influence the susceptibility of pluripotent stem cells to erosion, Cloutier and colleagues⁵³³ tested several *in vitro* culture conditions for human and mouse embryonic stem cells. Based on their results, they were able to re-focus the risk from the O₂ concentration, a widely assumed risk factor for the erosion of the X-chromosome⁴⁶⁷, to media composition. In their study, the use of mTeSR1 led to the rapid loss of the *XIST* cloud surrounding the inactive X-chromosome. They identified LiCl as the major factor responsible for the loss of *XIST* expression in hESCs, suggesting that this extracellular factor could influence the XCI process by inhibiting GSK-3 proteins⁵³⁴.

Before the publication of this article, mTeSR1 was routinely used in my laboratory for the cultivation of human iPSCs, and cells continued to be cultured in mTeSR1 for consistency in the study. The use of mTeSR1 can account for the rapid loss of the *XIST* cloud in two of the iPSC lines used in the study. However, although the mentioned study seemed to solve the molecular pathway leading to the loss of *XIST* expression due to LiCl activity, no data were available on the molecular mechanisms underlying this in hiPSCs. It remains unclear, why not all iPSC lines used in the PhD thesis exhibited erosion, despite being cultured in mTeSR1. These findings highlight the need for future studies to adopt culture media that protect against erosion mechanisms rather than contribute to their induction.

Several studies have reported heterogeneity in *XIST* expression among hPSC lines, reflecting the heterogeneous status of the inactive X-chromosome. Some lines maintained high *XIST*-expression during culturing, while others lost it early after reprogramming⁴⁷⁰. The complex range of erosion statuses of the inactive X-chromosome should be monitored by assessing X-linked genes and *XIST* expression, as well as by examining *XIST* distribution in the nucleus using RNA-FISH technology. By employing those approaches together, Patel and colleagues⁴⁶⁹ identified different statuses for the X-chromosome in hESCs derived from human blastocysts. They observed two main conditions for X-chromosome activity: embryonic stem cells with two active X-chromosomes (XaXa), and cells with one active and one inactive X-chromosomes (XaXi). The cells belonging to the second group were characterized by high heterogeneity in their status, with cells having an inactive X-chromosome coated by *XIST* (XaXi^{*XIST*}), and cells with intermediate levels of erosion, where the inactive X-chromosome had lost the *XIST* cloud but was still transcriptionally silenced (XaXi), or where the loss of *XIST* led to variable extents of erosion (XaXe). Interestingly, the X-inactivation status was conserved during hESC differentiation.

More recently, Raposo and colleagues⁴⁷⁰ investigated the erosion status of a set of hiPSCs. They identified similar heterogeneity in the status of the inactive X-chromosome across different clones analysed, with some clones maintaining a stable inactive X-chromosome coated by *XIST*, while others exhibited varying degrees of erosion, consistent with observations in the literature for hiPSCs^{535,536}. Unlike embryonic stem cells, which maintain an active X chromosome, hiPSCs undergo complete X-chromosome inactivation (XCI) due to their representation of a later embryonic stage, and no XaXa cells were found in the samples (). The Xi status present in the specific hiPS clone was also maintained during iPSC differentiation, confirming that no XCI could be induced during differentiation protocols and that the erosion is propagated also in differentiated cells, in line with previous studies⁴⁷⁴.

Following the literature, the characterization of the inactive X-chromosome status was performed on the iPSC lines used in my thesis. *XIST* expression was analysed using both RNA-seq and RT-qPCR data, and its nuclear localization was investigated at the single-cell level using RNA-FISH technology. Additionally, the presence of the Xi territory in the nucleus was assessed through H3K27me3 staining, a repressive histone mark accumulated on the inactive X-chromosome. These analyses revealed a

heterogenous status among the cell lines used in the study. The M- and J-lines were found to be eroded, showing neither *XIST* expression nor the presence of the Xi territory. In contrast, the A-line exhibited only very few cells eroding the Xi during extensive culturing. However, data derived from QUASEP analysis showed that the extent of erosion present in the M- and J-line iPSCs did not lead to biallelic expression of X-linked genes in iPSCs. For this reason, the results obtained from all three cell lines could be valuable for further assessing facultative escape genes in cell lines derived from different donors.

Despite the erosion affecting the global loss of the *XIST* cloud covering the inactive X-chromosome, only a subset of genes appears to be impacted^{464,466,470}. Specific regions along the eroded X-chromosome (Xe) have been identified in hPSCs as being subjected to gene reactivation by Vallot and colleagues^{466,537}. This confirms that erosion is not a chromosome-wide phenomenon, a finding also supported by Patel and colleagues⁴⁶⁹. Notably, the overall extent of biallelic expression from the eroded X-chromosome did not reach the chromosome-wide expression levels observed in XaXa hESCs, reinforcing the notion that erosion is a process affecting specific regions of the X-chromosome. Those results were further corroborated by a study published by Raposo and colleagues⁴⁷⁰, where they reported that in eroded hiPSCs, the Xp22 and Xq22-q23 regions were most affected by the absence of the *XIST* cloud, with gene expression observed upon erosion.

Additionally, the erosion status was reported to influence the extent of biallelic expression of escape genes. In non-eroded hiPSCs, differences in allele-specific expression levels were observed for certain escape genes, with lower expression from the Xi compared to the Xa⁴⁷⁰. However, the absence of *XIST* coating leads to the artificial upregulation of these escapees from the Xi, resulting in similar expression levels between the active and inactive X-chromosomes in hiPSCs. The relationship between escape gene expression levels and *XIST* presence was further explored in a recent pre-print by Haut and colleagues⁴⁶⁴. By modulating *XIST* expression levels in mouse clonal NPCs, they demonstrated that ectopic expression of *XIST* can influence the expression levels of escape genes in a dosage-dependent manner. The expression of escape genes from the inactive allele was reduced when *XIST* expression was increased. However, constitutive escape genes were less affected by *XIST* levels and still able to escape XCI even at high *XIST* expression levels. That was not the case for facultative escape genes, where fluctuations in *XIST* expression induced significant changes in their expression, resulting in complete silencing from the Xi at the highest *XIST* expression levels. The different susceptibility of constitutive and facultative escape genes to *XIST* levels may reflect distinct mechanisms regulating the expression from the Xi and suggest different evolutionary mechanisms underlying the phenomenon. However, the inducible overexpression of *XIST* in mouse NPCs might not reflect the situation *in vivo* in NPCs, so these results should be interpreted cautiously. The results obtained from the Haut study implicate that, in the artificial context of erosion of the inactive X-chromosome, the absence of *XIST* could lead to improper allele-specific expression ratios between the active and inactive X-chromosome, primarily affecting facultative escapees, and not reflecting the physiological gene expression from the inactive X-chromosome.

In my thesis, two lines with erosion hallmarks and one non-eroded iPSCs line were used to investigate the facultative escape from XCI during neuronal differentiation. The absence of *XIST* expression in the M-line and J-line could influence the extent of Xi expression of the identified facultative escape genes, the reactivated and late-silenced genes. No striking allele-specific expression differences between cell lines were observed for the late-silenced genes detected in more than one cell line, suggesting conserved expression levels of late-silenced genes from the inactive X-chromosome and no influence from the erosion status (Figure 18.F). On the contrary, the reactivated genes showed complex individual- and cell-type specific Xi expression levels. When examining the reactivated genes in both *XIST*-positive and *XIST*-negative cell lines, two of the reactivated candidates, *FMR1* and *HCFCL1*, detected reactivated in

the A- and M-line, and A- and J-line respectively, exhibited higher Xi expression in the *XIST*-negative lines (Figure 18.B). However, this was not the case for most of the other reactivated genes, where the Xi expression levels were similar to or higher in the *XIST*-positive line. This pattern was true for *MIDI*, *HUWE-1*, and *FTX*, genes detected as reactivated in all three cell lines and thus identified as strong reactivated candidates (Figure 18.B). These results suggest that expression of reactivated genes from the inactive X chromosome was not extensive in *XIST*-negative lines, suggesting that the identified facultative escape genes in the *XIST*-negative lines are not necessarily the result of erosion mechanisms.

To date, no effective method has been established to stably reverse the erosion status of hPSCs^{538,539}. In a study conducted by Motosugi and team, the erosion status of hPSCs was rescued by targeting the promoter region of the *XIST* gene using CRISPR-Cas9 technology. Two approaches were employed. The first one utilized homologous directed repair mechanism (HDR) to introduce a new copy of *XIST* into the genome. The second approach tested in the study involved targeting the promoter region of the *XIST* gene with a single guide RNA (gRNA) to trigger the non-homologous end-joining repair mechanism (NHEJ). It had been shown that the NHEJ could induce changes in the methylation status of targeted regions⁵⁴⁰. For this reason, NHEJ was employed to change the methylation profile of the *XIST* promoter, inducing the expression of the lncRNA in eroded, edited cells.

While this method shows promise in restoring *XIST* expression in eroded cells, its success depends on the efficacy of the NHEJ mechanism in inducing changes in the epigenetic status of the *XIST* regulatory region. A second approach used for restoring *XIST* expression and proper X-inactivation status was recently published by Agostinho de Sousa and colleagues (). They induced the conversion of eroded primed hPSCs into naïve hPSCs, which acquired *XIST* expression in a similar pattern to that of pre-implantation epiblast cells. *XIST* expression was maintained during the capacitation process, which converts naïve hPSCs back into a primed state. Unfortunately, the rescued hPSCs maintained proper XCI status for only a few passages, after which eroded cells began to accumulate.

In conclusion, no effective method has been developed to stably revert the erosion status of the X-chromosome so far. The extensive changes induced by erosion, leading to a global imbalance of X-linked gene expression and influencing the global proteome of the iPSCs⁵⁴¹, combined with the inability to reverse this status, mean that using eroded cell lines could drastically influence experimental outcomes. However, cell systems that recapitulate the mutation and specific genetic background of a patient are of fundamental importance for studying disease phenotypes. This was the case for the M-line, derived from a heterozygous patient with a mutation in the *MIDI* gene. The introduction of the same mutation into both eroded (J-line) and non-eroded (A-line) iPSCs helped understand and isolate the effects of the mutation in brain organoids derived from these lines. This revealed that the disease phenotype was associated specifically with the *MIDI* mutation and not with the eroded status. These data were generated in collaboration with Prof. Marisa Karow, Dr. Sven Falk, Dr. Rhadika Mennon, Elisa Gabassi, and Sarah Frank.

The relationship between reactivated genes and erosion was further investigated by tracking the allele-specific expression of two reactivated genes, *GPM6B* and *MIDI*, monitored using QUASEP assays and allele-specific RT-PCR in the eroded M- and J-lines and the non-eroded A-line. In all iPSC lines tested, these genes were strictly monoallelic, with no evidence of gene expression leakage from the eroded Xi. Immunostaining for H3K27me3 revealed the absence of this repressive epigenetic mark on the Xi in the nuclei of the M- and J-lines. While H3K27me3 predominantly accumulates on the long arm of the X-chromosome, the short arm is mainly marked by H3K9me3. *MIDI* and *GPM6B* may represent a special case of reactivated genes located in a region particularly resistant to erosion, which could explain their consistent monoallelic expression even in eroded iPSCs.

Further examination of *XIST* expression in two additional M-line clones previously characterised, M16 and M19 (Thesis, Dr. Stephan Kaseberg), revealed that the M16 NPCs, but not the M16 iPSCs, exhibited high level of *XIST* expression similar to that of fibroblasts. This difference could have resulted from using originally non-eroded iPSCs for the NPC differentiation. Allele-specific RT-PCR showed that *XIST*-positive M16 NPCs reactivated the *MIDI* allele on the inactive X-chromosome. These findings helped us to understand that *MIDI* reactivation is not linked to the global absence of repressive marks on the Xi induced by the erosion, but rather the result of a local mechanism triggered by the differentiation of iPSCs into NPCs, occurring in both *XIST*-positive and *XIST*-negative cells.

5.3 Identification of novel categories of facultative escapees

5.3.1 Technologies for the identification of escape genes

Since the discovery of escape genes, various approaches have been employed to identify genes expressed from the inactive X-chromosome, both *in vivo* and *in vitro*. Although the mouse model has been instrumental in uncovering new escape genes, it has not been effective in identifying human-specific escape genes⁵⁴². While some escape genes are conserved between humans and mice, the number of escape genes appears to be substantially higher in humans, suggesting the presence of evolutionarily unique mechanisms. Therefore, it is crucial to develop advanced technologies and strategies specifically tailored to study X-chromosome inactivation (XCI) and escape genes in humans.

One of the earliest technologies used to detect expression from the inactive X-chromosome (Xi) in humans involved mouse/human cell hybrids. To complement SNP-based allele-specific expression analysis conducted in fibroblasts with non-random XCI, Carrel and Willard used hybrid rodent/human somatic cells⁴⁶⁵. In this approach, human fibroblasts were fused together with murine cells, and the resulting hybrid cells containing either the human active (Xa) or inactive (Xi) X-chromosomes were isolated. The expression of any X-linked gene could be investigated by designing a human specific RT-PCR assay, allowing to study the expression of human X-linked genes^{465,531,543}. Through this method, Carrel and Willard assigned allele-specific expression status to 612 X-linked genes, identifying that 15% of those were constitutive escapees, while 10% were identified as facultative escapees. However, this technology has limitations, including the artificial context of the human Xi and the non-human environment⁵⁴⁴.

To overcome the limitations of earlier method and identify escapees in a more physiological context, researchers turned to human RNA-sequencing datasets. If a gene escapes from the Xi, its expression in females will be the sum of the expression from both the Xa and Xi, usually resulting in higher expression in females compared to males. Leveraging this concept, Oliva and colleagues⁴⁵⁹ identified a set of human tissue-specific facultative escape genes by analysing gene expression across 44 tissues from 838 individuals. By focusing on known escape genes, they confirmed that a set of previously reported escape genes exhibited a female expression bias. Furthermore, they identified 40 additional X-linked genes with a female expression bias that could potentially represent newly identified escape genes. However, this approach has its limitations, since it indirectly identifies escape genes. For instance, sex-biased expression could result from secondary sex traits, such as hormonal influences, or even from the impact of other escape genes on global gene expression. Additionally, escape genes do not always result in higher expression in females, and stringent thresholds for detecting female sex bias may mask the presence of escape genes with low expression from the inactive X-chromosome. Moreover, sex-specific expression differences could be attributed to the overall number of X-chromosomes, e.g. in individuals with gonosomal aneuploidies, rather than true sex-specific differences, further complicating this indirect approach and underscoring the need for more direct methods of identifying escape candidates^{545,546}.

Furthermore, epigenetic marks have been employed to identify escape genes. If a gene is actively transcribed from the Xi, it loses its repressive marks and acquires active chromatin marks. In the study of escape genes, a set of epigenetic marks have been found to correlate with escape status⁴⁸⁴. It was shown that genes subject to XCI were enriched with repressive histone marks, while genes escaping the XCI were associated with open chromatin marks and bound by RNA-Polymerases. By comparing active and inactive chromatin marks between males and females, researchers could identify candidate escape genes based on their epigenetic profiles. That approach was utilised by Balaton and Brown⁵⁴⁷, who compared the XCI status of genes with several histone marks and DNA methylation patterns using published data. They confirmed that genes subject to XCI were enriched with heterochromatin marks, while escape genes were associated with euchromatin marks. This data was also used to train a model capable of predicting whether a gene is subject to XCI based on its epigenetic status. The advantage of this method lies in its ability to annotate genes as escapees without relying on specific expression patterns in particular tissues^{548,549}. However, like the RNA-sequencing approach, this method also shares the limitation of only identifying correlations between epigenetic marks and escape status, without directly confirming whether a gene is truly escaping XCI.

To directly investigate expression from the Xi, allele-specific analysis tools have been employed. These analyses rely on the presence of genomic variants in X-chromosomal genes, which allow researchers to distinguish between expression from the active (Xa) and inactive (Xi) X-chromosomes. However, a significant limitation of this approach is that the samples used often come from *in vivo* material or primary cells, both of which are subject to random XCI process. This randomness results in a mosaic pattern of cells, with either one X-chromosome or the other being active, making it challenging to discern whether expression is originating from the Xa or Xi based on SNP analysis³⁰¹. To address this challenge, algorithms have been developed to analyse large datasets and predict allele-specific expression in non-skewed individuals⁵²⁹. While bioinformatics predictions and large datasets can help, the use of non-random XCI systems, where one X chromosome is preferentially inactivated, can facilitate the detection of allele-specific expression. However, this approach is not feasible with *in vivo* human samples, as only very few individuals naturally exhibit skewed X-chromosomal inactivation across multiple tissues^{532,550-552}. Furthermore, attempting to induce skewed XCI through genetic manipulation *in vitro* could introduce artificial changes, potentially compromising the integrity of the system^{444,553}. Finally, the availability of human primary material is limited, often restricted to primary cells or post-mortem tissues. This limitation not only hinders the acquisition of human samples for analysis but also restricts observations to specific time points in development, making it impossible to track facultative escape gene expression throughout the human developmental process.

Leveraging single-cell RNA-sequencing technologies offers a promising alternative to employing skewed XCI systems. At single-cell resolution, it is possible to determine monoallelic or biallelic expression of X-linked genes, even within populations with random XCI^{301,554}. However, several technical limitations make the use of this technology still preliminary in the escapee field. For accurate allele-specific analysis, a deep coverage of the entire RNA sequence is required to ensure that expression of the genomic variants is detected. While bulk RNA-seq approaches can achieve this level of coverage, only a few scRNA-seq methods allow full gene body sequencing. Furthermore, allele expression could be altered by the detection of transcriptional burst or the sequencing of cell duplets, requiring an elevated number of single cells sequenced to find consistent and reliable results.

In my thesis, RNA-sequencing was employed to identify escape genes in iPSCs, NPCs, and neurons. The RNA-sequencing experiments were performed in bulk, by comparing the overall gene expression between iPSCs, NPCs, and neurons from three different cell lines. Thanks to the deep sequencing performed, the X-linked gene expression could be followed at allele-specific resolution. Allelic

expression of X-linked genes was derived by the analysis of all heterozygous genomic variants located on the X-chromosome. If a SNP or insertion/deletion (indel) was found to be monoallelically expressed, it was assumed that the expression originated from the active X-chromosome or that the genomic variant was not heterozygous in our lines. However, if the genomic variant was found to be heterozygously expressed, it was inferred that one part of the expression was coming from the active X-chromosome, while the other part from the inactive X-chromosome. Furthermore, for the first time the allele-specific expression was tracked throughout neuronal differentiation. This approach allowed a better understanding of the trajectory of inactivation and escape during neuronal differentiation, showing changes in their expression depending on the cell type and stage of differentiation.

5.3.2 Constitutive and facultative escape gene expression across *in vitro* neural differentiation

Based on genomic variant calling of bulk RNA-sequencing data, we identified escape genes with a constant expression from the inactive allele across different cell lines and cell types, as well as genes with variable expression from the inactive X-chromosome depending on the differentiation status. For the first time, we described two novel categories of facultative escape genes characterized by changes in their expression from the inactive X-chromosome during neuronal differentiation. The first category includes genes that were monoallelically expressed in iPSCs but became biallelically expressed in NPCs and/or neurons, referred to as reactivated genes in the thesis. Conversely, we identified genes that were expressed from both the active and inactive X-chromosomes in iPSCs but only from the active X-chromosome in NPCs and neurons; these are referred to as late-silenced genes.

We detected a total of 81 genes escaping the XCI in iPSCs, NPCs, and neurons in a similar manner. Among these, only 16 genes consistently escaped in all three cell lines tested, confirming their status as constitutive escapees, steadily escaping the XCI across all cell lines and cell types. However, due to the nature of the analysis, the number of escape genes identified reflected the distribution of genetic variants in the genome, particularly in expressed gene regions. In our study we considered genes escaping in all three cell types in at least one cell line as full-escapees and regarded them as constitutive escapees, assuming that the absence of detection in the other cell lines was due to the lack of heterozygous expressed variants.

We reported 124 genes as facultative escapees, with their expression from the inactive X-chromosome varying depending on the cell types. Of these, 95 genes were assigned to the reactivated category, while 29 were late-silenced. The same limitations regarding genetic variant distribution applied for the detection of these categories. Follow-up analyses on the gene categories were based on genes escaping in one or two cell types of at least two cell lines, further increasing the confidence in the allele-specific analysis.

By extending the approximation to include all genes identified in at least one cell line and considering the total number of identified genes in each category, we observed percentages of detected constitutive and facultative escape genes matching the data reported in the literature. Considering recent estimates of 867 genes located on the X-chromosome⁵⁵⁵, we detected 9.3% of these as constitutive escapees and 14.3% as facultative escapees. These percentages align with literature reports, where the percentage of escape genes in humans is generally reported as more than 20%, with facultative escape genes outnumbering constitutive escape genes^{447,485}.

5.3.3 RNA-FISH nuclear signal distribution recapitulates the dynamic X-chromosomal gene expression

Further validation of the dynamic allele-specific expression of X-linked genes during neuronal differentiation was achieved using an RNA-FISH approach through RNAScope technology. Unlike traditional RNA-FISH methods, RNAScope includes amplification steps following probe hybridization and signal detection, which enables the accurate detection of small amounts of RNA in samples. Custom probes were designed to target intronic regions of X-linked genes. By employing low annealing temperatures between probes and samples, the probes could specifically hybridize with RNA, while DNA remained double stranded. This technique ensured that the intron-targeting probes bound exclusively to unspliced RNA found in the nucleus, termed pre-mRNA, which is still associated with the transcriptional site on the chromosome of origin. Coupling these custom probes with probes targeting the lncRNA *XIST* allowed the detection of allele-specific expression of X-linked target genes at single-nucleus resolution.

RNA-FISH has been extensively used and validated in studies aiming to elucidate allele-specific expression of X-linked genes⁵⁵⁶. Most studies employing RNA-FISH focus on human embryonic stem cells (hESCs), and use probes targeting the lncRNAs *XIST* and *XACT*, as well as probes for the pre-mRNA of X-linked transcripts, including both, constitutive escape genes and inactive genes. However, due to the non-clonal nature of hESCs and their pre-XCI status, tracking the allele-specific expression of X-linked genes during pluripotent stem cell differentiation poses significant challenges. In my thesis, I have innovatively employed RNA-FISH to track the allele-specific expression of X-linked facultative escape genes during 2D neural differentiation of a clonal population of pluripotent stem cells. This approach provides new insights into the facultative reactivation mechanism at single-nucleus resolution.

RNA-FISH technology was employed to examine the allele-specific signal of X-linked genes within the nuclei of iPSCs and NPCs. For the RNA-FISH experiments, we utilized *XIST*-positive iPS clones that exhibited a stable *XIST* cloud covering the Xi, primarily using the *XIST*-positive A-ctrl iPS line. The presence of a specific probe signal indicated the expression of a specific gene from either the active X-chromosome (Xa) or the inactive X-chromosome (Xi). Genes expressed from the Xa were identified by the absence of co-localization with the *XIST* cloud, whereas genes expressed from the Xi were detected by their co-localization with the *XIST* cloud.

The analysis of the nuclear signal distribution confirmed the observation that the expression levels of constitutive escape genes remained consistent throughout the neural differentiation process of iPSCs. For example, the allele-specific expression of the full-escapee *NLGN4X* was tracked during the differentiation from iPSCs to NPCs. The similar percentage of cells with a biallelic expression of *NLGN4X* in both the iPSC and NPCs samples corroborated the results of the RNA-seq analysis, which identified *NLGN4X* as a full-escapee gene across all three cell lines, with a similar allele expression ratio between the active and inactive X-chromosome.

Since *NLGN4X* is detected as full-escape gene in the RNA-seq analysis, a biallelic signal should be expected in all nuclei. However, while the RNA-seq provides detailed allele-specific RNA expression levels, measuring the gene expression from the active and inactive X-chromosomes, RNA-FISH technology was employed to quantify the percentage of nuclei showing monoallelic and biallelic signals, regardless of expression levels coming from the active and the inactive X-chromosome. Therefore, even if the proportion of nuclei with biallelic signals was not the majority but rather represented half of the total cell population, the signal coming from the inactive X-chromosome can still account for the overall expression levels detected in bulk RNA-seq.

Specific RNA-FISH probes were designed for two genes, *MIDI* and *GPM6B*, which were identified as reactivated in all three cell lines. Based on the analysis approach described above, we observed reactivation at a single-nucleus resolution. In iPSCs, the percentage of nuclei with biallelic signals for these genes was below 15%, while in NPCs, this percentage increased significantly, exceeding 30%. This increase in biallelic expression in NPCs confirms, at the single-nuclei level, the changes in allele-specific expression of these genes during neural differentiation. Despite the significant rise in biallelic nuclei in NPCs, the proportion did not reach the levels observed for *NLGN4X* in either iPSCs or NPCs. This discrepancy highlights the differences in allele-specific expression between full escapee genes and reactivated genes, reinforcing the observed variations in expression levels between constitutive and facultative escape genes in the RNA-seq data. Furthermore, although the percentage of NPCs biallelically expressing the reactivated *MIDI* and *GPM6B* genes represents only 25% to 30% of the total NPC population, they can still account for the Xi expression levels detected by allele-specific RNA-seq analysis, as observed for the *NLGN4X* gene.

In NPCs, only 30% of nuclei showed a biallelic signal of the reactivated genes, while 50% of nuclei showed a biallelic signal of the full escape gene *NLGN4X*. This might reflect a high degree of variability in the expression of constitutive and facultative escape genes from the Xi. This observed variability could be linked to the presence of different cell types within the NPC sample. Recent investigations on human NPCs, obtained following the same dual SMAD inhibition differentiation protocol, have shown that the obtained samples are heterogeneous. Qiu and colleagues⁵⁰⁷ differentiated hiPSCs into NPCs and performed scRNA-seq experiments. The analysis of the control cells revealed the presence of cell clusters with different identities. In addition to the expected neural progenitor cells and neural stem cells clusters, oligodendrocyte progenitor cells and radial glial progenitor cells were also found in the sample. The presence of different cell types indicates a heterogeneous cell population in the samples used for the RNA-FISH and RNA-seq analysis. On single nucleus resolution using RNA-FISH, the heterogeneous samples showed only a subset of cells reactivating the X-linked genes and biallelically expressing the full escape gene. This might reflect that the biallelic expression of escape genes is cell type specific, where reactivated genes are biallelically expressed only in certain cell types during neuronal development, and full escape genes are more broadly expressed across cell types. However, further experiments involving the cell type characterization of the samples at single cell resolution and the concomitant detection of reactivated and full escape gene expression should be employed to answer this exciting question.

Expression differences between constitutive and facultative escape genes were previously observed in the literature. Constitutive escape genes tend to maintain stable biallelic expression from both X-chromosomes⁴⁸⁵, which may be related to the presence of homologous genes on the Y chromosome, making the dosage compensation of those genes of fundamental importance⁵⁵⁷. Furthermore, the function of constitutive escapees is linked to essential cell biological processes, such as transcription, translation, and epigenetic modifications, making the dosage compensation of those genes essential⁵⁵⁷.

For instance, Hauth and colleagues⁴⁶⁴ demonstrated experimentally that while the expression levels of constitutive escape genes could be modulated by altering *XIST* expression in their in vitro cell system, these genes consistently managed to escape XCI. By contrast, by increasing *XIST* levels facultative escape genes could be silenced from the inactive X-chromosome, highlighting a dynamic response to XCI modulation. Differences in *XIST*-dependency between constitutive and facultative genes suggest that constitutive escapees may have evolved gene-specific mechanisms to bypass XCI repression, thereby not being subject to any of the repression present on the Xi. Conversely, facultative escape genes could be the result of a more flexible control, influenced by cell type specific and individual-specific factors that modulate XCI. This variability reflects regions along the Xi that are more prone to escape

the XCI when repression levels induced by *XIST* are lower than usual, controlled by a combination of different epigenetic and chromatin mechanisms⁴⁴⁷. Such dynamics are particularly relevant during development, where *XIST* levels fluctuate⁵⁵⁸. The ability of facultative escape genes, like the reactivated genes identified in our study, to gain escape status could confer significant advantages to developing female cells and tissues, potentially contributing to observed sex biases in disease susceptibility⁴³².

5.3.4 Cell type specific reactivation of genes from the inactive X-Chromosome

During the follow-up analysis of the identified escape genes, we focused on the reactivated gene category, recognizing their potentially crucial role during neural differentiation. Reactivated genes were the only category significantly enriched for genes associated with neurodevelopmental disorders (NDD), highlighting a strong correlation between their biallelic expression and their function in neuronal development. The modulation of NDD-associated gene expression during key stages of neurodevelopment may influence the phenotype manifestation of NDD in females, potentially explaining the sex bias observed in the literature for many NDDs.

Moreover, reactivated genes were found to be significantly closer to each other on the X-chromosome compared to other escape genes identified. It has been reported in the literature that one of the key distinctions between constitutive and facultative escape genes is that the latter often form topologically associating domain-like structures (TAD-like). These TAD-like structures interact at the 3D conformational level, facilitating the expression of genes located within and at the boundaries of these domains⁴⁴⁷. Hauth and colleagues⁴⁶⁴ demonstrated that the formation of TAD-like structures containing clusters of facultative escape genes could be dynamically influence by altering *XIST* levels. This might implicate that the 3D genome organization of facultative escape genes into transcriptionally active domains can be modulated, allowing for temporal- and cell type-specific regulation of these genes. By contrast, constitutive escape genes are reported to be excluded from *XIST* coating, not being subject to any epigenetic deposition of repressive marks nor part of dynamic TAD-like structures^{463,559-561}. Cluster analysis conducted in my thesis confirmed the tendency of reactivated genes to be located closer to each other on the X-chromosome than would be expected by random distribution, raising the intriguing possibility that this proximity may influence 3D chromatin conformation. Further investigation is needed to explore the chromatin state of the reactivated genes during neural differentiation, which could shed light on the mechanisms driving their coordinated expression and potential roles in neurodevelopment.

During pluripotent stem cell differentiation, the transcriptome reshapes to align gene expression with the specific needs of the developing cell type. This dynamic process ensures that genes are expressed in a manner conducive to the identity and function of the differentiated cells. However, genes escaping the dosage compensation from the Xi could lead to profound changes in the global transcriptome, particularly if they are reactivated during differentiation. The broader impact of the reactivated genes was indirectly measured by looking at the protein-protein interaction (PPI) map focusing on the interactions between the proteins encoded by reactivated genes and all proteins expressed in neural progenitor cells (NPCs) and neurons. The analysis of these interaction networks revealed that many of the reactivated genes play critical roles in fundamental cellular processes, including chromatin organization, RNA binding, and ubiquitination. These functions are essential for maintaining cellular integrity, regulating gene expression, and controlling protein degradation pathways. The cell type-specific reactivation of these genes, particularly during neurodevelopmental stages, will therefore induce widespread changes in the global transcriptome. For instance, alterations in chromatin-binding proteins could lead to changes in chromatin accessibility and gene expression profiles, while disruptions in RNA-binding proteins could affect RNA stability, splicing, or translation. Similarly, changes in ubiquitination processes could impact protein turnover and signal transduction pathways, further influencing cellular

function. These potential changes may contribute to the differences observed between male and female transcriptomes and interactomes in developing brain tissue, as reported in the literature⁵⁶². The sexual dimorphism of the brain, reflected in differences in structure, function, and susceptibility to certain neurodevelopmental disorders, could be (partially) driven by the differential expression of facultative escape genes between males and females. In this context, the reactivation of specific X-linked genes during neuronal differentiation could be a central factor contributing to the complex interplay of genetic, epigenetic, and environmental influences that shape brain development and function differently in males and females.

An intriguing aspect of the reactivated genes' expression pattern is that some genes escape the Xi only at the NPCs stage, others only in neurons, and some in both cell types. This suggests that these genes may have stage-specific functions. In mice it has been reported that specific genes escape from XCI in neural stem cells *in vitro*⁴⁶⁴, and *in vivo* in brain tissue⁴⁴⁸. In humans, however, hardly any studies have directly investigated escape genes in neuronal tissue. The ability to escape XCI in a specific cell type is likely linked to changes in the epigenetic state of the genes. From our intersection between the list of reactivated genes and those subject to XCI with the ENCODE data on the epigenetic landscape of human male and female fetal brain tissue, we observed that reactivated genes were mostly associated with active chromatin marks rather than those repressing transcription, significantly in female but not in male cells. However, these data did not fully resolve the complexity of the cellular landscape present in the brain, which was only partially observed with our bulk RNA-seq analysis. Recently, the development of the MUSIC technology by Wen and colleagues could shed light on the cellular differences in brain tissue⁵⁶³. Multinucleic acid interaction mapping in single cells (MUSIC) enables the concomitant profiling of multiplex chromatin and RNA interactions, along with gene expression profiling, at single-cell resolution. The newly developed technology was applied to analyse human brain samples from female donors, and cell type specific differences were observed in the binding of *XIST* to the X-chromosome. This means that different cell types within the brain showed differences in the interaction between *XIST* RNA and the chromatin of the X-chromosome, with some cell types such as excitatory neurons, having a greater difference in the interaction between *XIST* and the Xi within the cell population compared to inhibitory neurons and astrocytes. Furthermore, reduced binding of *XIST* to the Xi led to higher affinity in the 3D structure between Xa and Xi, suggesting that the absence of *XIST* binding could lead to open chromatin regions and gene expression on the Xi in a similar way as the genes located on the Xa. These results align with the Xi-specific modulation of chromatin conformation and gene expression induced by *XIST* levels observed by Hauth and colleagues and suggest that, *in vivo*, the reduced presence of *XIST* in specific neural cell types could lead to the expression of facultative escape genes.

The observed differences in reactivated escape gene expression in NPCs and neurons, indicating potential cell-type specific variations, were further supported by results from a collaborating research group involved in the project. Brain organoids were generated from the A-line and from another iPSC line characterized by high *XIST* expression. Single-nucleus RNA-seq was performed to compare brain organoids with iPSC samples, and the data were analysed at allele-specific resolution. The transcriptome analysis identified several neuronal lineage trajectories arising from the brain organoid samples, mimicking the complexity of the brain organogenesis *in vitro* (Figure 2.A). Biallelic expression of a subset of X-linked genes with traceable SNPs was followed at single nucleus resolution (Figure 2.B), revealing the presence of biallelic expression in some iPSC sample clusters, as well as in specific neural lineages, such as neural progenitor cells and dopaminergic neurons. Furthermore, the presence of *XIST* was confirmed in iPSCs and maintained during brain organoid differentiation, while the pluripotent-specific *XACT* transcript was expressed in iPSCs but not in neuronal cells (Figure 2.C). This approach allowed for a detailed examination of inactive X-chromosome usage in specific neural and glial cell lineages, confirming that the biallelic expression of certain reactivated genes could be cell type specific

(Figure 2). As observed in specific cell populations in vivo in the brain⁵⁶³, the chromatin conformation of the Xi might change and lead to differences in the cell type population itself. This is further supported by the snRNA-seq data, where differences in the biallelic expression along a single differentiation trajectory were observed.

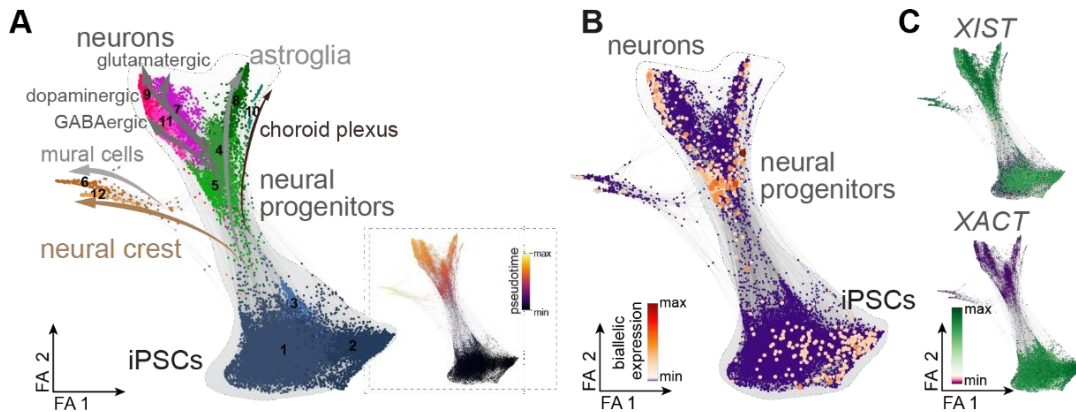


Figure 41. Dynamic biallelic expression of X-chromosomal genes across differentiation trajectories resolved on single cell level. A) Force-directed graph embedding of data received through snRNA-seq of iPSCs and organoids (d30) using the A-ctrl and the hoik1 iPSC lines. RNA-velocity based pseudotime is indicated in the dashed box. Number of cells total: 18.638, A-ctrl: 901, hoik1: 17.737. B) Degree of biallelic expression of X-linked genes projected onto the force-directed graph. C) Expression of *XIST* and *XACT* in snRNA-seq data.

5.3.5 Xp22.2 region: a hotspot for reactivated genes

Following the characterization of the reactivated gene category, we concentrated on a specific gene cluster for further validation analysis of the reactivated process. This cluster is located on the short arm of the X-chromosome, specifically in the Xp22.2 region. This region is part of Strata 4 and 5, segments of the X-chromosome that evolved later during evolution. After the mutation that led to the emergence of the *SRY* gene on the Y-chromosome, a series of inversions shaped the evolution of the X-chromosome, reducing the likelihood of recombination between the X and Y chromosomes^{564–568}. Strata 4 and 5 represent the most recent inversion events that occurred in eutherian mammals⁵⁶⁹.

Throughout the evolution of the sex chromosomes, some X-linked genes retained biallelic expression and remained unaffected by XCI due to the presence of homologous genes on the Y-chromosome. The presence of Y homology creates a critical need to balance gene expression between males and females, exerting evolutionary pressure on the gene copy located on the inactive X-chromosome to avoid silencing⁵⁵⁷. This is the case for X-linked genes that possess a Y-gametolog, as these genes constitutively escape XCI to balance their expression between males and females⁵⁶⁷. Another example is the escape genes located in the pseudoautosomal regions (PAR) 1 and 2, which still undergo homologous recombination during male meiosis and are equivalent between X and Y chromosomes. All genes located in the PAR1 region escape XCI, while only the distal genes located in the PAR2 region escape XCI, with the more proximal X-linked genes in the PAR2 being silenced on both X and Y chromosomes⁵⁷⁰. However, such dosage sensitivity is not a universal feature of all escape genes, as many escape genes have lost their Y homologs⁵⁷¹. Interestingly, escape genes are more commonly found on the short arm of the X-chromosome (Xp) than on the long one (Xq), with escape genes on the Xp forming clusters, while those on the Xq arm are found individually along the chromosome⁴⁸⁴. This may reflect the recent evolutionary addition of the Xp arm during the divergence between marsupials and eutherian

mammals⁴⁶⁵, along with the differences observed between the chromatin conformation of constitutive and facultative escape genes, with the latter forming cluster domains on the X-chromosome. Some of the facultative escape genes located on Xp may be influenced by signals coming from neighbouring escape genes, leading to their expression from the Xi^{447,572}.

Although we did not find a clear enrichment of reactivated and late-silenced genes on the short arm of the X-chromosome, we identified two reactivated genes located on the distal part of the Xp in close proximity to each other, indicating the presence of a reactivation cluster in that region.

Two genes located in the Xp22.2 region, *MIDI1* and *GPM6B*, were identified as reactivated in all three cell lines. The structure and function of the *MIDI1* gene were extensively described in the introduction. The *GPM6B* gene encodes the glycoprotein membrane 6B protein, a member of the proteolipid protein (PLP) family, which is highly expressed in the brain, particularly in astrocytes, neurons, and oligodendrocytes^{573–575}. The GPM6B protein plays a crucial role in axon myelination and development. Mutations in *GPM6B* are linked to psychiatric disorders, as demonstrated by the deletion of *GPM6B* in male mice using CRISPR-Cas9 technology^{573,576}. In humans, *GPM6B* expression was found to be significantly downregulated in the post-mortem brains of suicidal victims⁵⁷⁷. Interestingly, when examining gene expression relative to gender, *GPM6B* was the only gene showing female-biased expression, with greater downregulation observed in the brains of male suicidal victims compared to female⁵⁷⁷.

GPM6B and *MIDI1* reactivated genes were selected for additional confirmation analysis. They were detected as reactivated in three cell lines derived from three different female individuals. Through additional bulk and single-cell analyses, we confirmed that these genes were reactivated from the inactive X-chromosome in neuronal cells. Comparing these findings with recent escape studies in mice, *MIDI1* was annotated as a constitutive escapee, while *GPM6B* was categorized as a facultative escapee^{448,464}. In mice, *GPM6B* is classified as a brain-specific escape gene, suggesting a conserved role for this gene in neuronal cells between mice and humans⁴⁴⁸. In mice, *MIDI1* is located in the boundary region of the PseudoAutosomal Region (PAR) on the X-chromosome, an area known for escaping dosage compensation and behaving like an autosomal region. Several studies in mice have identified *MIDI1* as a constitutive escape gene, escaping XCI also in brain tissue^{578,579}. In humans, however, *MIDI1* is located in the Xp22.2 region, outside the PAR^{484,579}. This difference in chromosomal location might reflect variations in *MIDI1* expression and function between mice and humans, where *MIDI1* maintains escape properties in humans but does not constitutively escape. This suggests that *MIDI1* may have translocated from the PAR boundary region to Xp22.2, altering its escape properties, possibly influenced by neighbouring facultative escape genes. Further single-cell analyses are needed to determine whether *MIDI1* and *GPM6B* are reactivated together in the same cells or if the reactivation does not involve a reactivated domain, but rather single genes located in the Xp22.2. Given the proximity of these genes on the short arm of the X-chromosome, however, and the established fact that human escape genes often cluster along the X-chromosome, the reactivated cluster in the Xp22.2 region could well be biallelically expressed in the same cells during neuronal differentiation. Additional RNA-FISH experiments using pre-mRNA probes for both *MIDI1* and *GPM6B*, together with *XIST* probes, would be required to confirm this hypothesis.

5.3.6 Time and cell lineage specific reactivation of *MIDI1* influences the phenotype of OS syndrome

The confirmation that the *MIDI1* gene is reactivated during *in vitro* neuronal differentiation raises an intriguing question, whether this reactivation is present and could influence the disease phenotype associated with *MIDI1* mutations. Mutations in the *MIDI1* gene are associated with a monogenic

neurodevelopmental disorder known as Opitz-BBB/G (OS) syndrome. The M-line iPSCs possess a heterozygous 4 bp deletion in exon 9 of *MIDI* and were used to study the effect of this mutation in both 2D and 3D neuronal systems. Clones expressing the mutant allele and those expressing the wild-type allele were isolated during reprogramming. Previous experiments on brain organoids, as well as 2D NPCs and neurons, demonstrated that *MIDI* is reactivated in neuronal cells compared to iPSCs, confirming the reactivation observed in the RNA-seq data (Thesis, Dr. Kaseberg Stephan). To further investigate the role of the *MIDI* mutation in neuronal development, the same 4 base pair deletion was introduced into the exon 9 of *MIDI* gene in the J-line and A-line, creating both heterozygous and homozygous clones. Edited iPS cells were used to generate brain organoids and perform 2D neural differentiation. The patient-derived mutation served as a proxy to study the allele-specific expression of *MIDI*. Reactivation of the allele on the inactive X-chromosome was confirmed for both cell lines in NPCs and neurons using cells heterozygous for the *MIDI* mutation in my thesis. The data showed that in the presence of a heterozygous *MIDI* mutation, the inactive wild type allele could be dynamically reactivated in neural cells. The importance of *MIDI* reactivation is supported by literature, where it has been shown that MIDI1 protein can form heterodimers between mutant and wild type proteins^{388,580}. This suggests that in reactivating cells, wild type MIDI1 protein interacts with the mutant MIDI1 protein, potentially balancing the negative effects on ubiquitin and microtubule binding due to the 4 base pair mutation.

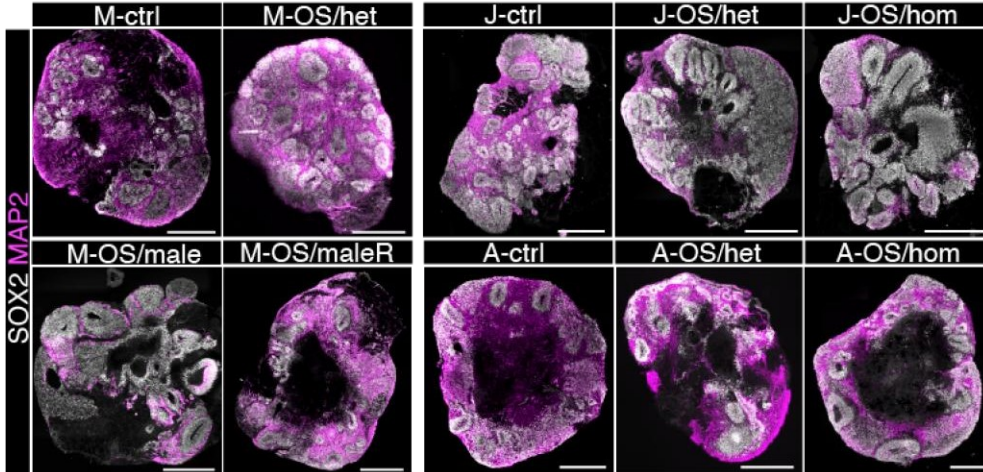
The use of the A- and J-OS heterozygous clones allowed us to deeply study *MIDI* reactivation in detail during the early stages of NPC differentiation, identifying specific time points during the NPC differentiation for *MIDI* reactivation: passage 4 for the A-line and passage 6 for the J-line. This suggests that expression from the inactive X-chromosome is not only induced by neural differentiation but is also finely tuned during neuronal lineage differentiation. This observation is consistent with data from single nucleus RNA-seq (Figure 2), which detected lineage-specific and temporal-specific biallelic expression. Additionally, the QUASEP assay for *MIDI* allele-specific expression in J-OS heterozygous cells showed that reactivation could be detected starting from passage 6 of the NPCs, with *MIDI* expression from the inactive X-chromosome increasing until passage 12, where the biallelic ratio of expression stabilised and was maintained at similar levels in neurons.

The time window of *MIDI* reactivation was further investigated in *XIST*-positive J2 NPCs using RNA-FISH. RNA-FISH was performed with probes targeting *MIDI* and *GPM6B*, in combination with the visualization of the *XIST* cloud, and analysed similarly to the A-line samples. By examining the percentage of nuclei with monoallelic or biallelic pre-mRNA signals for *MIDI*, a significant increase in the percentage of nuclei with a biallelic signal was detected in the J-line NPCs, similar to what was observed in the A-line. However, by looking at the distribution of the single NPC samples used for the RNA-FISH analysis, striking differences in the extent of reactivated cells were observed among the J2 NPC samples used. The RNA-FISH samples were derived from young J2 NPCs at passages 5, 6 and 9. Using early-passages NPCs allowed us to pinpoint the onset of gene reactivation from the inactive X-chromosome in J-line NPCs. As shown in the bar plot, J-line NPCs at passage 5 had 20% of nuclei with a biallelic *MIDI* signal, and this percentage increased with subsequent passages, reaching 50% at passage 9. This result confirms the findings from the allele-specific RT-PCR and QUASEP assays in the J-OS heterozygous line, demonstrating, at single-nuclei resolution, changes in the allele-specific expression of *MIDI* during NPC culturing and validating the time-specific reactivation of *MIDI*.

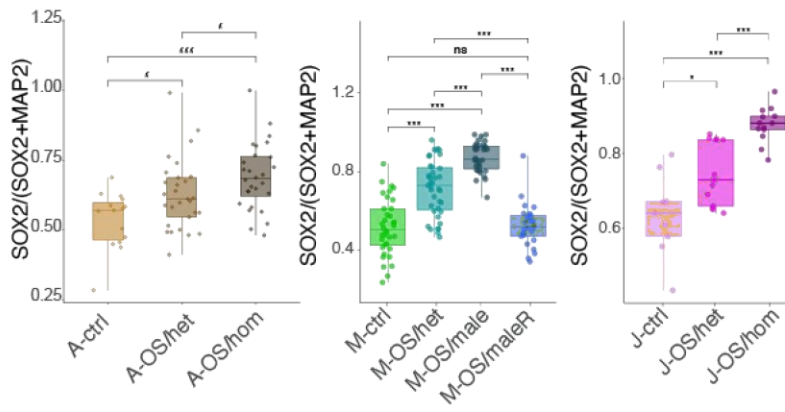
Brain organoids derived from M-, A-, and J-lines were analysed to understand the role of *MIDI* mutation and the impact of *MIDI* reactivation on the disease phenotype. In collaboration with a research group from Erlangen (Prof. Karow and Dr. Falk), brain organoids were analysed at day 30 of differentiation, and *MIDI* reactivation was detected via allele-specific RT-PCR (Figure 3.C, D, E). A clear phenotype

associated with a homozygous *MIDI* mutation in female or a hemizygous mutation in male brain organoids was observed (Figure 3.A, B). However, female brain organoids derived from heterozygous iPSCs clones expressing the mutant *MIDI* allele and reactivating the wild type allele exhibited an intermediate phenotype. This suggests a crucial effect of *MIDI* reactivation during neuronal differentiation, which can influence the phenotype associated with the NDD monogenic OS disease.

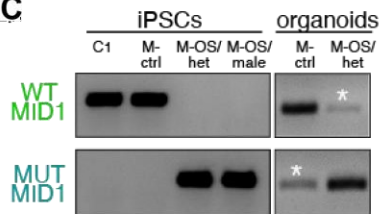
A



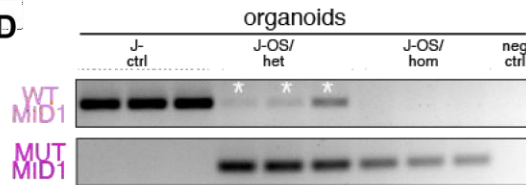
B



C



D



E

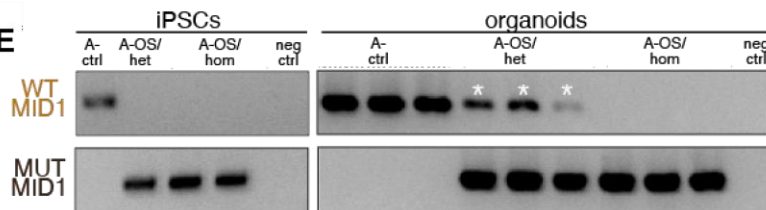


Figure 42. Dynamic biallelic expression of X-chromosomal genes in patient-derived iPSCs harboring heterozygous or homozygous *MID1* mutations. A) Immunostaining of organoid sections of the different experimental lines showing SOX2 (white) and MAP2 (magenta). Scale bars = 500 μm . B) Quantification of the fraction of SOX2⁺ area per total neural area (SOX2⁺ or MAP2⁺ area) in d30 organoids. Box- and jitter plot depicting the fold change of brain organoid area organized as VLZS compared to the respective ctrl line. A-ctrl, n=15; A-OS/het, n=29; A-OS/hom, n=28; exact P values (top to bottom): 0.035, 8×10^{-5} , 0.032. M-ctrl, n=36; M-OS/het, n=39; M-OS/male, n=34; M-OS/maleR, n=25; exact P values (top to bottom): 7.5×10^{-7} , 0.71, 1.4×10^{-13} , 2.8×10^{-7} , $< 2.2 \times 10^{-16}$, 1.3×10^{-7} . J-ctrl, n=9; J-OS/het, n=13; J-OS/hom, n=13; exact P values (top to bottom): 3.7×10^{-5} , 8.0×10^{-6} , 2.1×10^{-2} . Boxplots show median, quartiles (box), and range (whiskers), dots represent individual organoids; two-sided Wilcoxon rank sum test; *P<0.05, **P<0.01, ***P<0.001, ns P>0.05. C) Allele-specific RT-PCR revealing reactivation (*) of the inactive *MID1* allele in brain organoids derived from the M-iPSC lines. D) Reactivation (*) of the inactive *MID1* allele in brain organoids derived from the J-iPSCs as revealed by allele-specific RT-PCR. E) Reactivation (*) of the inactive *MID1* allele in brain organoids derived from the A-iPSCs as revealed by allele-specific RT-PCR.

In conclusion, the cell type- and time-specific reactivation of *MID1* impressively demonstrates the importance of the reactivation phenomenon during female neural development, both in the physiological and pathological context of female brain development, and sheds light into the molecular mechanisms underlying sexual dimorphism in neurodevelopmental disorders.

6 Conclusion

After successful characterization of human female iPSCs as a model of the early stages of human development, the clonal status of three iPSC lines was successfully confirmed by using QUASEP assay before differentiating them into NPCs and neurons. We took advantage of the skewed X-chromosome inactivation status of the female iPSC clones to establish an allele-specific RNA-seq analysis based on the detection of expressed heterozygous genomic variants. Two novel categories of facultative escape genes, reactivated and late-silenced genes, were discovered, paving the way that the gene expression from the inactive X-chromosome can dynamically change during neuronal differentiation. These data were further corroborated by single-nuclei RNA-Seq analysis, showing lineage-specific usage of the inactive X-chromosome.

Follow-up analyses focused on the reactivated genes because of their importance as human neuronal-specific facultative escape genes. We identified different epigenetic signatures of reactivated genes in female developing brain compared to male. Furthermore, disease-enrichment analysis showed an involvement of those genes in NDDs, and analysis of their function together with their interaction partners suggested important functions of reactivated genes during neurodevelopment, with a potential broad impact of the higher expression of those genes on the whole genome.

We choose two reactivated candidate genes, *MIDI* and *GPM6B*, detected reactivated in all three cell lines, for further validation of the RNA-seq analysis. Biallelic expression of the genes was detected by using QUASEP assay and allele-specific RT-PCR. Additionally, single-cell technology was employed for the confirmation of the reactivation process. RNA-FISH assay was established for the detection of the *MIDI* and *GPM6B* pre-mRNA together with the Xi territory, detected as *XIST* cloud. RNA-FISH results confirmed the differentiation dependent reactivation of *MIDI* and *GPM6B* in two different *XIST*-expressing iPSC clones.

MIDI reactivation was also further investigated in brain organoids derived from iPSCs carrying *MIDI* 4 base pair deletion variant. The phenotype analysis of brain organoids showed striking differences between control and affected homozygous females and hemizygous males, confirming the key role of *MIDI* during neuronal development. Surprisingly, the phenotype of female heterozygous brain organoids expressing the mutated *MIDI* allele and reactivating the wild type one showed an intermediate phenotype between the homozygous and the control female brain organoids. This suggested that the *MIDI* reactivation detected in heterozygous females influenced the brain organoids phenotype.

Taken together, the reactivation phenomenon seems associated with high dynamicity during neural differentiation, allowing changes in the gene expression from the Xi depending on the cell type and the differentiation status of the cell type population, profoundly influencing the transcriptome of the reactivating cells and potentially contributing to the resilience of female brain tissue against NDD.

7 References

1. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).
2. Raab, S., Klingenstein, M., Liebau, S. & Linta, L. A Comparative View on Human Somatic Cell Sources for iPSC Generation. *Stem Cells Int* **2014**, 768391 (2014).
3. Ray, A. *et al.* An Overview on Promising Somatic Cell Sources Utilized for the Efficient Generation of Induced Pluripotent Stem Cells. *Stem Cell Reviews and Reports* **2021** *17*:6 **17**, 1954–1974 (2021).
4. Hu, S. *et al.* Effects of cellular origin on differentiation of human induced pluripotent stem cell–derived endothelial cells. *JCI Insight* **1**, 85558 (2016).
5. Scesa, G., Adami, R. & Bottai, D. iPSC Preparation and Epigenetic Memory: Does the Tissue Origin Matter? *Cells* **2021**, Vol. 10, Page 1470 **10**, 1470 (2021).
6. Ebert, A. D. *et al.* Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* **2008** *457*:7227 **457**, 277–280 (2008).
7. Moretti, A. *et al.* Patient-Specific Induced Pluripotent Stem-Cell Models for Long-QT Syndrome. *New England Journal of Medicine* **363**, 1397–1409 (2010).
8. Park, I. H. *et al.* Disease-Specific Induced Pluripotent Stem Cells. *Cell* **134**, 877–886 (2008).
9. Maehr, R. *et al.* Generation of pluripotent stem cells from patients with type 1 diabetes. *Proc Natl Acad Sci U S A* **106**, 15768–15773 (2009).
10. Brennand, K. J. *et al.* Modelling schizophrenia using human induced pluripotent stem cells. *Nature* **2011** *473*:7346 **473**, 221–225 (2011).
11. Lee, G. *et al.* Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature* **2009** *461*:7262 **461**, 402–406 (2009).
12. Liu, G. H. *et al.* Modelling Fanconi anemia pathogenesis and therapeutics using integration-free patient-derived iPSCs. *Nature Communications* **2014** *5*:1 **5**, 1–17 (2014).
13. Rowe, R. G. & Daley, G. Q. Induced pluripotent stem cells in disease modelling and drug discovery. *Nature Reviews Genetics* **2019** *20*:7 **20**, 377–388 (2019).
14. Terhune, A. H., Bok, J., Sun, S. & Fu, J. Stem cell-based models of early mammalian development. *Development (Cambridge)* **149**, (2022).
15. Shoji, J. ya, Davis, R. P., Mummery, C. L. & Krauss, S. Global Literature Analysis of Organoid and Organ-on-Chip Research. *Adv Healthc Mater* **13**, 2301067 (2024).
16. Wu, X., Chen, Y., Kreutz, A., Silver, B. & Tokar, E. J. Pluripotent stem cells for target organ developmental toxicity testing. *Toxicological Sciences* **199**, 163–171 (2024).
17. Rispoli, P., Scanduzzi Piovesan, T., Decorti, G., Stocco, G. & Lucafò, M. iPSCs as a groundbreaking tool for the study of adverse drug reactions: A new avenue for personalized therapy. *WIREs Mechanisms of Disease* **16**, e1630 (2024).
18. Ma, X. *et al.* Cancer organoids: A platform in basic and translational research. *Genes Dis* **11**, 614–632 (2024).
19. Hendriks, D., Clevers, H. & Artigiani, B. CRISPR-Cas Tools and Their Application in Genetic Engineering of Human Stem Cells and Organoids. *Cell Stem Cell* **27**, 705–731 (2020).
20. Hendriks, D., Clevers, H. & Artigiani, B. CRISPR-Cas Tools and Their Application in Genetic Engineering of Human Stem Cells and Organoids. *Cell Stem Cell* **27**, 705–731 (2020).
21. Perlman, R. L. Mouse models of human disease: An evolutionary perspective. *Evol Med Public Health* **2016**, eow014 (2016).

22. Nakamura, T. *et al.* A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* 2016 537:7618 **537**, 57–62 (2016).
23. Guo, G. *et al.* Human naïve epiblast cells possess unrestricted lineage potential. *Cell Stem Cell* **28**, 1040 (2021).
24. Yanagida, A. *et al.* Naïve stem cell blastocyst model captures human embryo lineage segregation. *Cell Stem Cell* **28**, 1016-1022.e4 (2021).
25. Boroviak, T. & Nichols, J. Primate embryogenesis predicts the hallmarks of human naïve pluripotency. *Development* **144**, 175–186 (2017).
26. Ávila-González, D. *et al.* Pluripotent Stem Cells as a Model for Human Embryogenesis. *Cells* **12**, (2023).
27. Ávila-González, D. *et al.* Pluripotent Stem Cells as a Model for Human Embryogenesis. *Cells* **12**, (2023).
28. Zhu, M. *et al.* Human embryo polarization requires PLC signaling to mediate trophoderm specification. *Elife* **10**, (2021).
29. Rossant, J. & Tam, P. P. L. Early human embryonic development: Blastocyst formation to gastrulation. *Dev Cell* **57**, 152–165 (2022).
30. Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715–724 (2010).
31. Kuijk, E. W. *et al.* The roles of FGF and MAP kinase signaling in the segregation of the epiblast and hypoblast cell lineages in bovine and human embryos. *Development* **139**, 871–882 (2012).
32. Roode, M. *et al.* Human hypoblast formation is not dependent on FGF signalling. *Dev Biol* **361**, 358 (2012).
33. Meistermann, D. *et al.* Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification. *Cell Stem Cell* **28**, 1625-1640.e6 (2021).
34. Sutherland, A. Mechanisms of implantation in the mouse: Differentiation and functional importance of trophoblast giant cell behavior. *Dev Biol* **258**, 241–251 (2003).
35. Riveiro, A. R. & Brickman, J. M. From pluripotency to totipotency: an experimentalist’s guide to cellular potency. *Development* **147**, (2020).
36. Oldak, B. *et al.* Complete human day 14 post-implantation embryo models from naïve ES cells. *Nature* 2023 622:7983 **622**, 562–573 (2023).
37. Weatherbee, B. A. T. *et al.* Pluripotent stem cell-derived model of the post-implantation human embryo. *Nature* 2023 622:7983 **622**, 584–593 (2023).
38. Boroviak, T. & Nichols, J. Primate embryogenesis predicts the hallmarks of human naïve pluripotency. *Development* **144**, 175–186 (2017).
39. Tyser, R. C. V. *et al.* Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* 2021 600:7888 **600**, 285–289 (2021).
40. Meistermann, D. *et al.* Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification. *Cell Stem Cell* **28**, 1625-1640.e6 (2021).
41. Tyser, R. C. V. *et al.* Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* 2021 600:7888 **600**, 285–289 (2021).
42. Thomson, J. A. *et al.* Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science (1979)* **282**, 1145–1147 (1998).
43. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences* **78**, 7634–7638 (1981).
44. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).

45. Thomson, J. A. *et al.* Pluripotent cell lines derived from common marmoset (*Callithrix jacchus*) blastocysts. *Biol Reprod* **55**, 254–259 (1996).
46. Thomson, J. A. *et al.* Isolation of a primate embryonic stem cell line. *Proc Natl Acad Sci U S A* **92**, 7844–7848 (1995).
47. Thomson, J. A. *et al.* Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science (1979)* **282**, 1145–1147 (1998).
48. Zhou, J., Hu, J., Wang, Y. & Gao, S. Induction and application of human naive pluripotency. *Cell Rep* **42**, (2023).
49. Zhang, S. C., Wernig, M., Duncan, I. D., Brüstle, O. & Thomson, J. A. In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nature Biotechnology* *2001 19:12* **19**, 1129–1133 (2001).
50. Reubinoff, B. E. *et al.* Neural progenitors from human embryonic stem cells. *Nat Biotechnol* **19**, 1134–1140 (2001).
51. Kaufman, D. S., Hanson, E. T., Lewis, R. L., Auerbach, R. & Thomson, J. A. Hematopoietic colony-forming cells derived from human embryonic stem cells. *Proc Natl Acad Sci U S A* **98**, 10716–10721 (2001).
52. Steinbock, B. Life Before Birth: The Moral and Legal Status of Embryos and Fetuses. *Life Before Birth: The Moral and Legal Status of Embryos and Fetuses* 1–328 (2011) doi:10.1093/ACPROF:OSO/9780195341621.001.0001.
53. Eiges, R. *et al.* Developmental study of fragile X syndrome using human embryonic stem cells derived from preimplantation genetically diagnosed embryos. *Cell Stem Cell* **1**, 568–577 (2007).
54. Tulpule, A. *et al.* Knockdown of Fanconi anemia genes in human embryonic stem cells reveals early developmental defects in the hematopoietic lineage. *Blood* **115**, 3453–3462 (2010).
55. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences* **78**, 7634–7638 (1981).
56. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
57. Hanna, J. *et al.* Metastable Pluripotent States in NOD Mouse Derived ES Cells. *Cell Stem Cell* **4**, 513 (2009).
58. Boroviak, T., Loos, R., Bertone, P., Smith, A. & Nichols, J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat Cell Biol* **16**, 513–525 (2014).
59. Silva, J. *et al.* Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition. *PLoS Biol* **6**, e253 (2008).
60. Theunissen, T. W. *et al.* Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471–487 (2014).
61. Brons, I. G. M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* *2007 448:7150* **448**, 191–195 (2007).
62. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* *2007 448:7150* **448**, 196–199 (2007).
63. Takashima, Y. *et al.* Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254–1269 (2014).
64. Theunissen, T. W. *et al.* Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* **19**, 502 (2016).
65. de Sousa, J. A. *et al.* Epigenetic dynamics during capacitation of naïve human pluripotent stem cells. *Sci Adv* **9**, (2023).
66. Manor, Y. S., Massarwa, R. & Hanna, J. H. Establishing the human naïve pluripotent state. *Curr Opin Genet Dev* **34**, 35–45 (2015).
67. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* *2008 453:7194* **453**, 519–523 (2008).
68. Dong, C. *et al.* Derivation of trophoblast stem cells from naïve human pluripotent stem cells. *Elife* **9**, (2020).

69. Io, S. *et al.* Capturing human trophoblast development with naive pluripotent stem cells in vitro. *Cell Stem Cell* **28**, 1023-1039.e13 (2021).
70. Guo, G. *et al.* Human naive epiblast cells possess unrestricted lineage potential. *Cell Stem Cell* **28**, 1040 (2021).
71. Linneberg-Agerholm, M. *et al.* Naïve human pluripotent stem cells respond to Wnt, Nodal and LIF signalling to produce expandable naïve extra-embryonic endoderm. *Development* **146**, (2019).
72. Brons, I. G. M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **2007 448**:7150 **448**, 191–195 (2007).
73. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **2008 453**:7194 **453**, 519–523 (2008).
74. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **2007 448**:7150 **448**, 196–199 (2007).
75. Yu, L. *et al.* Blastocyst-like structures generated from human pluripotent stem cells. *Nature* **2021 591**:7851 **591**, 620–626 (2021).
76. Khan, S. A. & Theunissen, T. W. Modeling X-chromosome inactivation and reactivation during human development. *Curr Opin Genet Dev* **82**, (2023).
77. Kelley, K. W. & Paşca, S. P. Human brain organogenesis: Toward a cellular understanding of development and disease. *Cell* **185**, 42–61 (2022).
78. Bystron, I., Blakemore, C. & Rakic, P. Development of the human cerebral cortex: Boulder Committee revisited. *Nat Rev Neurosci* **9**, 110–122 (2008).
79. Sauer, F. C. Mitosis in the neural tube. *Journal of Comparative Neurology* **62**, 377–405 (1935).
80. Rakic, P. Specification of Cerebral Cortical Areas. *Science* (1979) **241**, 170–176 (1988).
81. Sidman, R. L. & Rakic, P. Neuronal migration, with special reference to developing human brain: a review. *Brain Res* **62**, 1–35 (1973).
82. Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011).
83. Bystron, I., Blakemore, C. & Rakic, P. Development of the human cerebral cortex: Boulder Committee revisited. *Nat Rev Neurosci* **9**, 110–122 (2008).
84. Smart, I. H. M. Proliferative characteristics of the ependymal layer during the early development of the mouse neocortex: a pilot study based on recording the number, location and plane of cleavage of mitotic figures. *J Anat* **116**, 67 (1973).
85. Miller, D. J., Bhaduri, A., Sestan, N. & Kriegstein, A. Shared and derived features of cellular diversity in the human cerebral cortex. *Curr Opin Neurobiol* **56**, 117–124 (2019).
86. Haubensak, W., Attardo, A., Denk, W. & Huttner, W. B. Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis. *Proc Natl Acad Sci U S A* **101**, 3196–3201 (2004).
87. Noctor, S. C., Martinez-Cerdeño, V., Ivic, L. & Kriegstein, A. R. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat Neurosci* **7**, 136–144 (2004).
88. Hansen, D. V., Lui, J. H., Parker, P. R. L. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
89. Fietz, S. A. *et al.* OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nat Neurosci* **13**, 690–699 (2010).
90. Hansen, D. V., Lui, J. H., Parker, P. R. L. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
91. Dehay, C., Kennedy, H. & Kosik, K. S. The Outer Subventricular Zone and Primate-Specific Cortical Complexification. *Neuron* **85**, 683–694 (2015).

92. Betizeau, M. *et al.* Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. *Neuron* **80**, 442–457 (2013).
93. Fietz, S. A. *et al.* OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nat Neurosci* **13**, 690–699 (2010).
94. Bae, B. Il, Jayaraman, D. & Walsh, C. A. Genetic Changes Shaping the Human Brain. *Dev Cell* **32**, 423 (2015).
95. Dehay, C., Kennedy, H. & Kosik, K. S. The Outer Subventricular Zone and Primate-Specific Cortical Complexification. *Neuron* **85**, 683–694 (2015).
96. Betizeau, M. *et al.* Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. *Neuron* **80**, 442–457 (2013).
97. Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011).
98. Kelley, K. W. & Paşca, S. P. Human brain organogenesis: Toward a cellular understanding of development and disease. *Cell* **185**, 42–61 (2022).
99. Nowakowski, T. J., Pollen, A. A., Sandoval-Espinosa, C. & Kriegstein, A. R. Transformation of the Radial Glia Scaffold Demarcates Two Stages of Human Cerebral Cortex Development. *Neuron* **91**, 1219–1227 (2016).
100. Hansen, D. V. *et al.* Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nat Neurosci* **16**, 1576–1587 (2013).
101. Herculano-Houzel, S. Coordinated Scaling of Cortical and Cerebellar Numbers of Neurons. *Front Neuroanat* **4**, (2010).
102. Haldipur, P. *et al.* Spatiotemporal expansion of primary progenitor zones in the developing human cerebellum. *Science* **366**, 454–460 (2019).
103. Florio, M. & Huttner, W. B. Neural progenitors, neurogenesis and the evolution of the neocortex. *Development* **141**, 2182–2194 (2014).
104. Liu, J. *et al.* The Primate-Specific Gene TMEM14B Marks Outer Radial Glia Cells and Promotes Cortical Expansion and Folding. *Cell Stem Cell* **21**, 635–649.e8 (2017).
105. Pollen, A. A. *et al.* Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).
106. Camp, J. G. *et al.* Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc Natl Acad Sci U S A* **112**, 15672–15677 (2015).
107. Johnson, M. B. *et al.* Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat Neurosci* **18**, 637–646 (2015).
108. Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science (1979)* **358**, 1318–1323 (2017).
109. Zhong, S. *et al.* A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
110. Polioudakis, D. *et al.* A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785–801.e8 (2019).
111. Eze, U. C., Bhaduri, A., Haeussler, M., Nowakowski, T. J. & Kriegstein, A. R. Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nature Neuroscience* **24**, 584–594 (2021).
112. Zhong, S. *et al.* Decoding the development of the human hippocampus. *Nature* **577**, 531–536 (2020).
113. Bocchi, V. D. *et al.* The coding and long noncoding single-cell atlas of the developing human fetal striatum. *Science* **372**, (2021).

114. Aldinger, K. A. *et al.* Spatial and cell type transcriptional landscape of human cerebellar development. *Nat Neurosci* **24**, 1163–1175 (2021).
115. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science* **370**, (2020).
116. La Manno, G. *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167**, 566–580.e19 (2016).
117. Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science (1979)* **358**, 1318–1323 (2017).
118. Zhong, S. *et al.* A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**:7697 **555**, 524–528 (2018).
119. Polioudakis, D. *et al.* A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785–801.e8 (2019).
120. Zhang, R., Quan, H., Wang, Y. & Luo, F. Neurogenesis in primates versus rodents and the value of non-human primate models. *Natl Sci Rev* **10**, (2023).
121. Benito-Kwiecinski, S. *et al.* An early cell shape transition drives evolutionary expansion of the human forebrain. *Cell* **184**, 2084–2102.e19 (2021).
122. Benito-Kwiecinski, S. *et al.* An early cell shape transition drives evolutionary expansion of the human forebrain. *Cell* **184**, 2084–2102.e19 (2021).
123. Reillo, I., De Juan Romero, C., García-Cabezas, M. Á. & Borrell, V. A role for intermediate radial glia in the tangential expansion of the mammalian cerebral cortex. *Cereb Cortex* **21**, 1674–1694 (2011).
124. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron* **89**, 248–268 (2016).
125. Carletti, B. & Rossi, F. Neurogenesis in the Cerebellum. <http://dx.doi.org/10.1177/1073858407304629> **14**, 91–100 (2007).
126. Vanderhaeghen, P. & Polleux, F. Developmental mechanisms underlying the evolution of human cortical circuits. *Nat Rev Neurosci* **24**, 213–232 (2023).
127. Chédotal, A. Roles of axon guidance molecules in neuronal wiring in the developing spinal cord. *Nat Rev Neurosci* **20**, 380–396 (2019).
128. Aida, T. & Feng, G. The dawn of non-human primate models for neurodevelopmental disorders. *Curr Opin Genet Dev* **65**, 160–168 (2020).
129. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron* **89**, 248–268 (2016).
130. Bystron, I., Rakic, P., Molnár, Z. & Blakemore, C. The first neurons of the human cerebral cortex. *Nat Neurosci* **9**, 880–886 (2006).
131. García-Moreno, F., López-Mascaraque, L. & De Carlos, J. A. Origins and migratory routes of murine Cajal-Retzius cells. *J Comp Neurol* **500**, 419–432 (2007).
132. Iacopetti, P. *et al.* Expression of the antiproliferative gene TIS21 at the onset of neurogenesis identifies single neuroepithelial cells that switch from proliferative to neuron-generating division. *Proc Natl Acad Sci U S A* **96**, 4639 (1999).
133. Chenn, A. & McConnell, S. K. Cleavage orientation and the asymmetric inheritance of Notch1 immunoreactivity in mammalian neurogenesis. *Cell* **82**, 631–641 (1995).
134. Marín-Padilla, M. The mammalian neocortex new pyramidal neuron: a new conception. *Front Neuroanat* **7**, (2013).
135. Takahashi, T., Goto, T., Miyama, S., Nowakowski, R. S. & Caviness, V. S. Sequence of neuron origin and neocortical laminar fate: relation to cell cycle of origin in the developing murine cerebral wall. *J Neurosci* **19**, 10357–10371 (1999).

136. Li, J. *et al.* Integrative genomic analysis of early neurogenesis reveals a temporal genetic program for differentiation and specification of preplate and Cajal-Retzius neurons. *PLoS Genet* **17**, e1009355 (2021).
137. Rakic, P. Prenatal development of the visual system in rhesus monkey. *Philos Trans R Soc Lond B Biol Sci* **278**, 245–260 (1977).
138. Hoerder-Suabedissen, A. & Molnár, Z. Development, evolution and pathology of neocortical subplate neurons. *Nat Rev Neurosci* **16**, 133–146 (2015).
139. Kanold, P. O. & Luhmann, H. J. The subplate and early cortical circuits. *Annu Rev Neurosci* **33**, 23–48 (2010).
140. Molnár, Z., Luhmann, H. J. & Kanold, P. O. Transient cortical circuits match spontaneous and sensory-driven activity during development. *Science* **370**, (2020).
141. Duque, A., Krsnik, Z., Kostović, I. & Rakic, P. Secondary expansion of the transient subplate zone in the developing cerebrum of human and nonhuman primates. *Proc Natl Acad Sci U S A* **113**, 9892–9897 (2016).
142. Cadwell, C. R., Bhaduri, A., Mostajo-Radji, M. A., Keefe, M. G. & Nowakowski, T. J. Development and Arealization of the Cerebral Cortex. *Neuron* **103**, 980–1004 (2019).
143. Balaram, P., Young, N. A. & Kaas, J. H. Histological features of layers and sublayers in cortical visual areas V1 and V2 of chimpanzees, macaque monkeys, and humans. *Eye Brain* **2014**, 5–18 (2014).
144. Telley, L. & Jabaudon, D. A mixed model of neuronal diversity. *Nature* **555**, 452–454 (2018).
145. Jakovcevski, I., Filipovic, R., Mo, Z., Rakic, S. & Zecevic, N. Oligodendrocyte development and the onset of myelination in the human fetal brain. *Front Neuroanat* **3**, (2009).
146. Yeung, M. S. Y. *et al.* Dynamics of oligodendrocyte generation and myelination in the human brain. *Cell* **159**, 766–774 (2014).
147. Zhu, Y. *et al.* Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, (2018).
148. Azevedo, F. A. C. *et al.* Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol* **513**, 532–541 (2009).
149. Rilling, J. K. *et al.* The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience* **2008** *11:4* **11**, 426–428 (2008).
150. Sousa, A. M. M., Meyer, K. A., Santpere, G., Gulden, F. O. & Sestan, N. Evolution of the Human Nervous System Function, Structure, and Development. *Cell* **170**, 226–247 (2017).
151. Rash, B. G. *et al.* Gliogenesis in the outer subventricular zone promotes enlargement and gyrification of the primate cerebrum. *Proc Natl Acad Sci U S A* **116**, 7089–7094 (2019).
152. Huang, W. *et al.* Origins and Proliferative States of Human Oligodendrocyte Precursor Cells. *Cell* **182**, 594–608.e11 (2020).
153. Menassa, D. A. & Gomez-Nicola, D. Microglial dynamics during human brain development. *Front Immunol* **9**, 370661 (2018).
154. Thion, M. S., Ginhoux, F. & Garel, S. Microglia and early brain development: An intimate journey. *Science* (1979) **362**, 185–189 (2018).
155. Bian, Z. *et al.* Deciphering human macrophage development at single-cell resolution. *Nature* **582**, 571–576 (2020).
156. Kracht, L. *et al.* Human fetal microglia acquire homeostatic immune-sensing properties early in development. *Science* (1979) **369**, 530–537 (2020).
157. Li, Q. *et al.* Developmental Heterogeneity of Microglia and Brain Myeloid Cells Revealed by Deep Single-Cell RNA Sequencing. *Neuron* **101**, 207–223.e10 (2019).
158. Hammond, T. R. *et al.* Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. *Immunity* **50**, 253–271.e6 (2019).

159. Peter R., H. Synaptic density in human frontal cortex — Developmental changes and effects of aging. *Brain Res* **163**, 195–205 (1979).
160. Koenderink, M. J. T., Uylings, H. B. M. & Mrzljak, L. Postnatal maturation of the layer III pyramidal neurons in the human prefrontal cortex: a quantitative Golgi analysis. *Brain Res* **653**, 173–182 (1994).
161. Huttenlocher, P.R., and Dabholkar, A.S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* **387**, 167–178. - Cerca con Google. [https://www.google.com/search?q=Huttenlocher%2CP.R.%2Cand+Dabholkar%2CA.S.+\(1997\).+Regional+differences+in+synaptogenesis+in+human+cerebral+cortex.+J.+Comp.+Neurol.+387%2C+167-178.&rlz=1C1GCEB_enDE988DE988&oq=Huttenlocher%2CP.R.%2Cand+Dabholkar%2CA.S.+\(1997\).+Regional+differences+in+synaptogenesis+in+human+cerebral+cortex.+J.+Comp.+Neurol.+387%2C+167-178.&gs_lcrp=EgZjaHJvbWUqBggAEEUYOzIGCAAQRrg70gEHODY3ajBqNKgCALACAQ&sourceid=chrome&ie=UTF-8](https://www.google.com/search?q=Huttenlocher%2CP.R.%2Cand+Dabholkar%2CA.S.+(1997).+Regional+differences+in+synaptogenesis+in+human+cerebral+cortex.+J.+Comp.+Neurol.+387%2C+167-178.&rlz=1C1GCEB_enDE988DE988&oq=Huttenlocher%2CP.R.%2Cand+Dabholkar%2CA.S.+(1997).+Regional+differences+in+synaptogenesis+in+human+cerebral+cortex.+J.+Comp.+Neurol.+387%2C+167-178.&gs_lcrp=EgZjaHJvbWUqBggAEEUYOzIGCAAQRrg70gEHODY3ajBqNKgCALACAQ&sourceid=chrome&ie=UTF-8)
162. Zhu, Y. *et al.* Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, (2018).
163. Ghosh, A., Antonini, A., McConnell, S. K. & Shatz, C. J. Requirement for subplate neurons in the formation of thalamocortical connections. *Nature* **347**, 179–181 (1990).
164. Molnár, Z., Luhmann, H. J. & Kanold, P. O. Transient cortical circuits match spontaneous and sensory-driven activity during development. *Science* **370**, (2020).
165. Ohtaka-Maruyama, C. *et al.* Synaptic transmission from subplate neurons controls radial migration of neocortical neurons. *Science (1979)* **360**, 313–317 (2018).
166. Huttenlocher, P.R., and Dabholkar, A.S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* **387**, 167–178. - Cerca con Google.
167. Peter R., H. Synaptic density in human frontal cortex — Developmental changes and effects of aging. *Brain Res* **163**, 195–205 (1979).
168. Sasai, Y. Cytosystems dynamics in self-organization of tissue architecture. *Nature 2013 493:7432* **493**, 318–326 (2013).
169. Clevers, H. Modeling Development and Disease with Organoids. *Cell* **165**, 1586–1597 (2016).
170. Shahbazi, M. N. & Zernicka-Goetz, M. Deconstructing and reconstructing the mouse and human early embryo. *Nature Cell Biology 2018 20:8* **20**, 878–887 (2018).
171. Pas, S. P. The rise of three-dimensional human brain cultures. *Nature 2018 553:7689* **553**, 437–445 (2018).
172. DeLong, G. R. Histogenesis of fetal mouse isocortex and hippocampus in reaggregating cell cultures. *Dev Biol* **22**, 563–583 (1970).
173. Crain, S. M., Alfei, L. & Peterson, E. R. Neuromuscular transmission in cultures of adult human and rodent skeletal muscle after innervation in vitro by fetal rodent spinal cord. *J Neurobiol* **1**, 471–489 (1969).
174. Garber, B. B. & Moscona, A. A. Reconstruction of brain tissue from cell suspensions. 2. Specific enhancement of aggregation of embryonic cerebral cells by supernatant from homologous cell cultures. *Dev Biol* **27**, 235–243 (1972).
175. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).
176. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnology 2009 27:3* **27**, 275–280 (2009).
177. Mertens, J., Marchetto, M. C., Bardy, C. & Gage, F. H. Evaluating cell reprogramming, differentiation and conversion technologies in neuroscience. *Nature Reviews Neuroscience 2016 17:7* **17**, 424–437 (2016).
178. Wang, C. *et al.* Scalable Production of iPSC-Derived Human Neurons to Identify Tau-Lowering Compounds by High-Content Screening. *Stem Cell Reports* **9**, 1221–1233 (2017).
179. Pas, S. P. The rise of three-dimensional human brain cultures. *Nature 2018 553:7689* **553**, 437–445 (2018).

180. Sasai, Y. Cytosystems dynamics in self-organization of tissue architecture. *Nature* 2013 493:7432 **493**, 318–326 (2013).
181. Eiraku, M. *et al.* Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* 2011 472:7341 **472**, 51–56 (2011).
182. Camp, J. G. *et al.* Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc Natl Acad Sci U S A* **112**, 15672–15677 (2015).
183. Mariani, J. *et al.* FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* **162**, 375–390 (2015).
184. Kadoshima, T. *et al.* Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc Natl Acad Sci U S A* **110**, 20284–20289 (2013).
185. Pasca, A. M. *et al.* Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nature Methods* 2015 12:7 **12**, 671–678 (2015).
186. Yoon, S. J. *et al.* Reliability of human cortical organoid generation. *Nature Methods* 2018 16:1 **16**, 75–78 (2018).
187. Sloan, S. A. *et al.* Human Astrocyte Maturation Captured in 3D Cerebral Cortical Spheroids Derived from Pluripotent Stem Cells. *Neuron* **95**, 779–790.e6 (2017).
188. Trevino, A. E. *et al.* Chromatin accessibility dynamics in a model of human forebrain development. *Science* (1979) **367**, (2020).
189. Gordon, A. *et al.* Long-term maturation of human cortical organoids matches key early postnatal transitions. *Nature Neuroscience* 2021 24:3 **24**, 331–342 (2021).
190. Pasca, A. M. *et al.* Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nature Methods* 2015 12:7 **12**, 671–678 (2015).
191. Parsch, J. & Ellegren, H. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet* **14**, 83–87 (2013).
192. Graves, J. A. M. Sex chromosome specialization and degeneration in mammals. *Cell* **124**, 901–914 (2006).
193. Graves, J. A. M., Koina, E. & Sankovic, N. How the gene content of human sex chromosomes evolved. *Curr Opin Genet Dev* **16**, 219–224 (2006).
194. Bachtrog, D. Y chromosome evolution: emerging insights into processes of Y chromosome degeneration. *Nat Rev Genet* **14**, 113 (2013).
195. Hughes, J. F. & Page, D. C. The Biology and Evolution of Mammalian Y Chromosomes. *Annu Rev Genet* **49**, 507–527 (2015).
196. Veyrunes, F. *et al.* Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res* **18**, 965 (2008).
197. Studies in spermatogenesis : Stevens, N. M. (Nettie Maria), 1861-1912 : Free Download, Borrow, and Streaming : Internet Archive. <https://archive.org/details/studiesinspermat01stev/page/n3/mode/2up>.
198. X-chromosome Dosage Compensation. [https://www.queensu.ca/academia/forsdyke/xchromos.htm#EVIDENCE OF THE PRECISION OF GENETIC ADAPTATION](https://www.queensu.ca/academia/forsdyke/xchromos.htm#EVIDENCE_OF_THE_PRECISION_OF_GENETIC_ADAPTATION).
199. Barr, M. L. & Bertram, E. G. A Morphological Distinction between Neurones of the Male and Female, and the Behaviour of the Nucleolar Satellite during Accelerated Nucleoprotein Synthesis. *Nature* 1949 163:4148 **163**, 676–677 (1949).
200. Allocyly of the X-Chromosome in Tumors and Normal Tissues* | Cancer Research | American Association for Cancer Research. <https://aacrjournals.org/cancerres/article/20/4/541/474395/Allocyly-of-the-X-Chromosome-in-Tumors-and-Normal>.
201. Lyon, M. F. Gene action in the X-chromosome of the mouse (*mus musculus* L.). *Nature* (1961) doi:10.1038/190372a0.

202. Ohno, S. Sex Chromosomes and Sex-Linked Genes. **1**, (1967).
203. Williams, T. M. & Carroll, S. B. Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nat Rev Genet* **10**, 797–804 (2009).
204. Parsch, J. & Ellegren, H. The evolutionary causes and consequences of sex-biased gene expression. *Nat Rev Genet* **14**, 83–87 (2013).
205. Lindenfors, P., Gittleman, J. L. & Jones, K. E. Sexual size dimorphism in mammals. *Sex, Size and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism* (2007) doi:10.1093/ACPROF:OSO/9780199208784.003.0003.
206. Ralls, K. Mammals in which females are larger than males. *Q Rev Biol* **51**, 245–276 (1976).
207. Tombak, K. J., Hex, S. B. S. W. & Rubenstein, D. I. New estimates indicate that males are not larger than females in most mammal species. *Nature Communications* 2024 15:1 **15**, 1–7 (2024).
208. Wells, J. C. K. Sexual dimorphism of body composition. *Best Pract Res Clin Endocrinol Metab* **21**, 415–430 (2007).
209. Callewaert, F. *et al.* Sexual dimorphism in cortical bone size and strength but not density is determined by independent and time-specific actions of sex steroids and IGF-1: evidence from pubertal mouse models. *J Bone Miner Res* **25**, 617–626 (2010).
210. Chen, S. *et al.* Hearing the physical condition: The relationship between sexually dimorphic vocal traits and underlying physiology. *Front Psychol* **13**, 983688 (2022).
211. Naqvi, S. *et al.* Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science (1979)* **365**, (2019).
212. Aguet, F. *et al.* The impact of sex on gene expression across human tissues. *Science (1979)* **369**, (2020).
213. Rodríguez-Montes, L. *et al.* Sex-biased gene expression across mammalian organ development and evolution. *Science (1979)* **382**, (2023).
214. Rodríguez-Montes, L. *et al.* Sex-biased gene expression across mammalian organ development and evolution. *Science (1979)* **382**, (2023).
215. Hu, R., McDonough, A. A. & Layton, A. T. Sex differences in solute and water handling in the human kidney: Modeling and functional implications. *iScience* **24**, (2021).
216. Mauvais-Jarvis, F. *et al.* Sex and gender: modifiers of health, disease, and medicine. *The Lancet* **396**, 565–582 (2020).
217. Wilkinson, N. M., Chen, H. C., Lechner, M. G. & Su, M. A. Sex Differences in Immunity. *Annu Rev Immunol* **40**, 75 (2022).
218. Eliot, L., Ahmed, A., Khan, H. & Patel, J. Dump the “dimorphism”: Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neurosci Biobehav Rev* **125**, 667–697 (2021).
219. Galupa, R. & Heard, E. X-chromosome inactivation: A crossroads between chromosome architecture and gene regulation. *Annual Review of Genetics* vol. 52 535–566 Preprint at <https://doi.org/10.1146/annurev-genet-120116-024611> (2018).
220. Whitworth, D. J. & Pask, A. J. The X factor: X chromosome dosage compensation in the evolutionarily divergent monotremes and marsupials. *Semin Cell Dev Biol* **56**, 117–121 (2016).
221. Mahadevaiah, S. K., Sangrithi, M. N., Hirota, T. & Turner, J. M. A. A single-cell transcriptome atlas of marsupial embryogenesis and X inactivation. *Nature* **586**, 612–617 (2020).
222. Grant, J. *et al.* Rxs is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature* **487**, 254–258 (2012).
223. Okamoto, I. *et al.* Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472**, 370–374 (2011).
224. Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* (2016) doi:10.1016/j.cell.2016.03.023.

225. Okamoto, I. *et al.* The X chromosome dosage compensation program during the development of cynomolgus monkeys. *Science* **374**, (2021).
226. Kobayashi, S. *et al.* Live imaging of X chromosome reactivation dynamics in early mouse development can discriminate naïve from primed pluripotent stem cells. *Development* **143**, 2958–2964 (2016).
227. Okamoto, I. *et al.* Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472**, 370–374 (2011).
228. Zhou, F. *et al.* Reconstituting the transcriptome and DNA methylome landscapes of human implantation. *Nature* **572**:7771 **572**, 660–664 (2019).
229. Okamoto, I. *et al.* The X chromosome dosage compensation program during the development of cynomolgus monkeys. *Science* **374**, (2021).
230. Tan, K. *et al.* Impaired imprinted X chromosome inactivation is responsible for the skewed sex ratio following in vitro fertilization. *Proc Natl Acad Sci U S A* **113**, 3197–3202 (2016).
231. Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev* **11**, 156–166 (1997).
232. Mugford, J. W., Yee, D. & Magnuson, T. Failure of extra-embryonic progenitor maintenance in the absence of dosage compensation. *Development* **139**, 2130–2138 (2012).
233. Borsani, G. *et al.* Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**, 325–329 (1991).
234. Brown, C. J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527–542 (1992).
235. Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* (1991) doi:10.1038/349038a0.
236. Heard, E., Mongelard, F., Arnaud, D. & Avner, P. Xist yeast artificial chromosome transgenes function as X-inactivation centers only in multicopy arrays and not as single copies. *Mol Cell Biol* **19**, 3156–3166 (1999).
237. Brockdorff, N. *et al.* The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515–526 (1992).
238. Clemson, C. M., McNeil, J. A., Willard, H. F. & Lawrence, J. B. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol* **132**, 259–275 (1996).
239. Brown, C. J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527–542 (1992).
240. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, (2013).
241. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137 (1996).
242. Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol Cell* **5**, 695–705 (2000).
243. Popova, B. C., Tada, T., Takagi, N., Brockdorff, N. & Nesterova, T. B. Attenuated spread of X-inactivation in an X;autosome translocation. *Proc Natl Acad Sci U S A* **103**, 7706–7711 (2006).
244. Loda, A. *et al.* Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. *Nat Commun* **8**, (2017).
245. Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev* **11**, 156–166 (1997).
246. Brockdorff, N. *et al.* Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* **351**, 329–331 (1991).

247. Loda, A., Collombet, S. & Heard, E. Gene regulation in time and space during X-chromosome inactivation. *Nat Rev Mol Cell Biol* **23**, 231–249 (2022).
248. Nesterova, T. B. *et al.* Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res* **11**, 833–849 (2001).
249. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* **30**, 167–174 (2002).
250. Hoki, Y. *et al.* A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development* **136**, 139–146 (2009).
251. Sakata, Y. *et al.* Defects in dosage compensation impact global gene regulation in the mouse trophoblast. *Development* **144**, 2784–2797 (2017).
252. Shi, Y. *et al.* Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev* **15**, 1140 (2001).
253. You, S. H. *et al.* Nuclear receptor co-repressors are required for the histone-deacetylase activity of HDAC3 in vivo. *Nat Struct Mol Biol* **20**, 182–187 (2013).
254. Guenther, M. G., Barak, O. & Lazar, M. A. The SMRT and N-CoR Corepressors Are Activating Cofactors for Histone Deacetylase 3. *Mol Cell Biol* **21**, 6091–6101 (2001).
255. Mikami, S. *et al.* Structural insights into the recruitment of SMRT by the corepressor SHARP under phosphorylative regulation. *Structure* **22**, 35–46 (2014).
256. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
257. Żylicz, J. J. *et al.* The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell* (2019) doi:10.1016/j.cell.2018.11.041.
258. Dossin, F. *et al.* SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature* **578**, 455–460 (2020).
259. Markaki, Y. *et al.* Xist nucleates local protein gradients to propagate silencing across the X chromosome. *Cell* **184**, 6174–6192.e32 (2021).
260. Pintacuda, G. *et al.* hnRNP K Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol Cell* **68**, 955–969.e10 (2017).
261. Nozawa, R. S. *et al.* Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nat Struct Mol Biol* **20**, 566–573 (2013).
262. Bousard, A. *et al.* The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. *EMBO Rep* **20**, (2019).
263. Almeida, M. *et al.* PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* **356**, 1081–1084 (2017).
264. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
265. Jansz, N. *et al.* Smchd1 Targeting to the Inactive X Is Dependent on the Xist-HnrnpK-PRC1 Pathway. *Cell Rep* **25**, 1912–1923.e9 (2018).
266. Bousard, A. *et al.* Exploring the role of Polycomb recruitment in Xist-mediated silencing of the X chromosome in ES cells. *bioRxiv* 495739 (2018) doi:10.1101/495739.
267. Beletskii, A., Hong, Y. K., Pehrson, J., Egholm, M. & Strauss, W. M. PNA interference mapping demonstrates functional domains in the noncoding RNA Xist. *Proc Natl Acad Sci U S A* **98**, 9215 (2001).
268. Sarma, K., Lévasséur, P., Aristarkhov, A. & Lee, J. T. Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome. *Proc Natl Acad Sci U S A* **107**, 22196–22201 (2010).

269. Yamada, N. *et al.* Xist Exon 7 Contributes to the Stable Localization of Xist RNA on the Inactive X-Chromosome. *PLoS Genet* **11**, (2015).
270. Sunwoo, H., Colognori, D., Froberg, J. E., Jeon, Y. & Lee, J. T. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc Natl Acad Sci U S A* **114**, 10654–10659 (2017).
271. Ridings-Figueroa, R. *et al.* The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev* **31**, 876–888 (2017).
272. Jeon, Y. & Lee, J. T. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**, 119–133 (2011).
273. Colognori, D., Sunwoo, H., Kriz, A. J., Wang, C. Y. & Lee, J. T. Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X. *Mol Cell* **74**, 101–117.e10 (2019).
274. Hasegawa, Y. *et al.* The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell* **19**, 469–476 (2010).
275. Kelsey, A. D. *et al.* Impact of flanking chromosomal sequences on localization and silencing by the human non-coding RNA XIST. *Genome Biol* (2015) doi:10.1186/s13059-015-0774-2.
276. Dixon-McDougall, T. & Brown, C. J. Multiple distinct domains of human XIST are required to coordinate gene silencing and subsequent heterochromatin formation. *Epigenetics Chromatin* **15**, 1–18 (2022).
277. Smeets, D. *et al.* Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. *Epigenetics & Chromatin* **7**, 1–27 (2014).
278. 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).
279. Roignant, J. Y. & Soller, M. m6A in mRNA: An Ancient Mechanism for Fine-Tuning Gene Expression. *Trends Genet* **33**, 380–390 (2017).
280. Patil, D. P. *et al.* m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
281. Loda, A. & Heard, E. Xist RNA in action: Past, present, and future. *PLoS Genet* **15**, (2019).
282. Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D. & Heard, E. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 644–649 (2004).
283. Mak, W. *et al.* Reactivation of the paternal X chromosome in early mouse embryos. *Science* **303**, 666–669 (2004).
284. Patrat, C. *et al.* Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. *Proc Natl Acad Sci U S A* **106**, 5198–5203 (2009).
285. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
286. Borensztein, M. *et al.* Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. *Nat Struct Mol Biol* **24**, 226–233 (2017).
287. Vallot, C. *et al.* XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. *Cell Stem Cell* (2017) doi:10.1016/j.stem.2016.10.014.
288. De Andrade E Sousa, L. B. *et al.* Kinetics of Xist-induced gene silencing can be predicted from combinations of epigenetic and genomic features. *Genome Res* **29**, 1087–1099 (2019).
289. Sahakyan, A. *et al.* Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. *Cell Stem Cell* **20**, 87–101 (2017).
290. Alfeghaly, C. *et al.* XIST dampens X chromosome activity in a SPEN-dependent manner during early human development. *Nature Structural & Molecular Biology* **2024** 1–12 (2024) doi:10.1038/s41594-024-01325-3.

291. Motosugi, N. *et al.* Deletion of lncRNA XACT does not change expression dosage of X-linked genes, but affects differentiation potential in hPSCs. *Cell Rep* **35**, (2021).
292. Carrel, L. & Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400–404 (2005).
293. Balaton, B. P. & Brown, C. J. Escape Artists of the X Chromosome. *Trends in Genetics* Preprint at <https://doi.org/10.1016/j.tig.2016.03.007> (2016).
294. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).
295. Balaton, B. P., Fornes, O., Wasserman, W. W. & Brown, C. J. Cross-species examination of X-chromosome inactivation highlights domains of escape from silencing. *Epigenetics Chromatin* **14**, 1–17 (2021).
296. Carrel, L. & Brown, C. J. When the Lyon(ized chromosome) roars: ongoing expression from an inactive X chromosome. *Philos Trans R Soc Lond B Biol Sci* **372**, (2017).
297. Clemson, C. M., Chow, J. C., Brown, C. J. & Lawrence, J. B. Stabilization and Localization of Xist RNA are Controlled by Separate Mechanisms and are Not Sufficient for X Inactivation. *J Cell Biol* **142**, 13 (1998).
298. Willard, H. F., Brown, C. J., Carrel, L., Hendrich, B. & Miller, A. P. Epigenetic and Chromosomal Control of Gene Expression: Molecular and Genetic Analysis of X Chromosome Inactivation. *Cold Spring Harb Symp Quant Biol* **58**, 315–322 (1993).
299. Navarro-Cobos, M. J., Balaton, B. P. & Brown, C. J. Genes that escape from X-chromosome inactivation: Potential contributors to Klinefelter syndrome. *Am J Med Genet C Semin Med Genet* **184**, 226–238 (2020).
300. Nesterova, T. B. *et al.* Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nat Commun* **10**, (2019).
301. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).
302. Garieri, M. *et al.* Extensive cellular heterogeneity of X inactivation revealed by single-cell allele-specific expression in human fibroblasts. *Proc Natl Acad Sci U S A* **115**, 13015–13020 (2018).
303. Balaton, B. P. & Brown, C. J. Contribution of genetic and epigenetic changes to escape from X-chromosome inactivation. *Epigenetics Chromatin* **14**, (2021).
304. Cotton, A. M. *et al.* Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Hum Mol Genet* **24**, 1528–1539 (2015).
305. Gershman, A. *et al.* Epigenetic patterns in a complete human genome. *Science* **376**, (2022).
306. Kucera, K. S. *et al.* Allele-specific distribution of RNA polymerase II on female X chromosomes. *Hum Mol Genet* **20**, 3964–3973 (2011).
307. Qu, K. *et al.* Individuality and variation of personal regulomes in primary human T cells. *Cell Syst* **1**, 51–61 (2015).
308. Patel, S. *et al.* Human Embryonic Stem Cells Do Not Change Their X Inactivation Status during Differentiation. *Cell Rep* **18**, 54–67 (2017).
309. Gravholt, C. H. *et al.* The Changing Face of Turner Syndrome. *Endocr Rev* **44**, 33–69 (2023).
310. Fukami, M., Seki, A. & Ogata, T. SHOX Haploinsufficiency as a Cause of Syndromic and Nonsyndromic Short Stature. *Mol Syndromol* **7**, 3 (2016).
311. Ottesen, A. M. *et al.* Increased Number of Sex Chromosomes Affects Height in a Nonlinear Fashion: A Study of 305 Patients With Sex Chromosome Aneuploidy. *Am J Med Genet A* **152A**, 1206 (2010).
312. Schwartz, C. E. *et al.* X-Linked intellectual disability update 2022. *Am J Med Genet A* **191**, 144–159 (2023).
313. Migeon, B. R. X-linked diseases: susceptible females. *Genetics in Medicine* **22**, 1156–1174 (2020).
314. Gonçalves, T. F. *et al.* KDM5C mutational screening among males with intellectual disability suggestive of X-Linked inheritance and review of the literature. *Eur J Med Genet* **57**, 138–144 (2014).

315. Lederer, D. *et al.* Deletion of KDM6A, a Histone Demethylase Interacting with MLL2, in Three Patients with Kabuki Syndrome. *Am J Hum Genet* **90**, 119 (2012).
316. Mo, J. *et al.* DDX3X: structure, physiologic functions and cancer. *Mol Cancer* **20**, (2021).
317. Khan, O. M. *et al.* The deubiquitinase USP9X regulates FBW7 stability and suppresses colorectal cancer. *J Clin Invest* **128**, 1326–1337 (2018).
318. Dunford, A. *et al.* Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. *Nat Genet* **49**, 10–16 (2017).
319. Ricketts, C. J. & Marston Linehan, W. Gender Specific Mutation Incidence and Survival Associations in Clear Cell Renal Cell Carcinoma (CCRCC). *PLoS One* **10**, e0140257 (2015).
320. Van Der Meulen, J. *et al.* Lymphoid Neoplasia: The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. *Blood* **125**, 13 (2015).
321. Scully, E. P., Haverfield, J., Ursin, R. L., Tannenbaum, C. & Klein, S. L. Considering how biological sex impacts immune responses and COVID-19 outcomes. *Nat Rev Immunol* **20**, 442–447 (2020).
322. Forsyth, K. S., Jiwrajka, N., Lovell, C. D., Toothacre, N. E. & Anguera, M. C. The conneXion between sex and immune responses. *Nat Rev Immunol* **24**, 487–502 (2024).
323. Souyris, M. *et al.* TLR7 escapes X chromosome inactivation in immune cells. *Sci Immunol* **3**, (2018).
324. Yu, B. *et al.* B cell-specific XIST complex enforces X-inactivation and restrains atypical B cells. *Cell* **184**, 1790–1803.e17 (2021).
325. Sibley, M. H. *et al.* Late-Onset ADHD Reconsidered With Comprehensive Repeated Assessments Between Ages 10 and 25. *Am J Psychiatry* **175**, 140–149 (2018).
326. Agnew-Blais, J. & Arseneault, L. Late-Onset ADHD: Case Closed or Open Question? *Am J Psychiatry* **175**, 481–482 (2018).
327. Asherson, P. & Agnew-Blais, J. Annual Research Review: Does late-onset attention-deficit/hyperactivity disorder exist? *J Child Psychol Psychiatry* **60**, 333–352 (2019).
328. Santos, S., Ferreira, H., Martins, J., Gonçalves, J. & Castelo-Branco, M. Male sex bias in early and late onset neurodevelopmental disorders: Shared aspects and differences in Autism Spectrum Disorder, Attention Deficit/hyperactivity Disorder, and Schizophrenia. *Neurosci Biobehav Rev* **135**, (2022).
329. Thapar, A., Cooper, M. & Rutter, M. Neurodevelopmental disorders. *Lancet Psychiatry* **4**, 339–346 (2017).
330. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. *Diagnostic and Statistical Manual of Mental Disorders* (2022) doi:10.1176/APPI.BOOKS.9780890425787.
331. Ismail, F. Y. & Shapiro, B. K. What are neurodevelopmental disorders? *Curr Opin Neurol* **32**, 611–616 (2019).
332. Bölte, S. *et al.* Standardised assessment of functioning in ADHD: consensus on the ICF Core Sets for ADHD. *Eur Child Adolesc Psychiatry* **27**, 1261–1281 (2018).
333. Bölte, S. *et al.* The Gestalt of functioning in autism spectrum disorder: Results of the international conference to develop final consensus International Classification of Functioning, Disability and Health core sets. *Autism* **23**, 449–467 (2019).
334. D’Souza, H. & Karmiloff-Smith, A. Neurodevelopmental disorders. *Wiley Interdiscip Rev Cogn Sci* **8**, (2017).
335. Morris-Rosendahl, D. J. & Crocq, M. A. Neurodevelopmental disorders-the history and future of a diagnostic concept. *Dialogues Clin Neurosci* **22**, 65–72 (2020).
336. Rapoport, J. L., Giedd, J. N. & Gogtay, N. Neurodevelopmental model of schizophrenia: update 2012. *Mol Psychiatry* **17**, 1228 (2012).
337. 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).

338. Faraone, S. V. & Larsson, H. Genetics of attention deficit hyperactivity disorder. *Mol Psychiatry* **24**, 562–575 (2019).
339. Bölte, S., Girdler, S. & Marschik, P. B. The contribution of environmental exposure to the etiology of autism spectrum disorder. *Cell Mol Life Sci* **76**, 1275–1297 (2019).
340. Mandy, W. & Lai, M. C. Annual Research Review: The role of the environment in the developmental psychopathology of autism spectrum condition. *J Child Psychol Psychiatry* **57**, 271–292 (2016).
341. Kiser, D. P., Rivero, O. & Lesch, K. P. Annual research review: The (epi)genetics of neurodevelopmental disorders in the era of whole-genome sequencing--unveiling the dark matter. *J Child Psychol Psychiatry* **56**, 278–295 (2015).
342. Bauer, G. R. Sex and Gender Multidimensionality in Epidemiologic Research. *Am J Epidemiol* **192**, 122–132 (2023).
343. Mauvais-Jarvis, F. *et al.* Sex and gender: modifiers of health, disease, and medicine. *The Lancet* **396**, 565–582 (2020).
344. Bhargava, A. *et al.* Considering Sex as a Biological Variable in Basic and Clinical Studies: An Endocrine Society Scientific Statement. *Endocr Rev* **42**, 219–258 (2021).
345. Cost, K. T. *et al.* Checking assumptions: Advancing the analysis of sex and gender in human health and psychological sciences. (2022) doi:10.31219/OSF.IO/C29KG.
346. Khrantsova, E. A., Davis, L. K. & Stranger, B. E. The role of sex in the genomics of human complex traits. *Nat Rev Genet* **20**, 173–190 (2019).
347. Tahira, A. C. *et al.* Putative contributions of the sex chromosome proteins SOX3 and SRY to neurodevelopmental disorders. *Am J Med Genet* **180**, 390 (2019).
348. Arambula, S. E. & McCarthy, M. M. Neuroendocrine-Immune Crosstalk Shapes Sex-Specific Brain Development. *Endocrinology* **161**, (2020).
349. Joel, D. Beyond the binary: Rethinking sex and the brain. *Neurosci Biobehav Rev* **122**, 165–175 (2021).
350. DeCasien, A. R., Guma, E., Liu, S. & Raznahan, A. Sex differences in the human brain: a roadmap for more careful analysis and interpretation of a biological reality. *Biology of Sex Differences* **2022 13:1** **13**, 1–21 (2022).
351. Ostatníková, D., Lakatošová, S., Babková, J., Hodosy, J. & Celec, P. Testosterone and the brain: from cognition to autism. *Physiol Res* **69**, S403–S419 (2020).
352. Zablotzky, B. *et al.* Prevalence and Trends of Developmental Disabilities among Children in the United States: 2009–2017. *Pediatrics* **144**, (2019).
353. Posserud, M. B., Skretting Solberg, B., Engeland, A., Haavik, J. & Klungsoyr, K. Male to female ratios in autism spectrum disorders by age, intellectual disability and attention-deficit/hyperactivity disorder. *Acta Psychiatr Scand* **144**, 635–646 (2021).
354. Faraone, S. V. *et al.* The World Federation of ADHD International Consensus Statement: 208 Evidence-based conclusions about the disorder. *Neurosci Biobehav Rev* **128**, 789–818 (2021).
355. Rutherford, M. *et al.* Gender ratio in a clinical population sample, age of diagnosis and duration of assessment in children and adults with autism spectrum disorder. *Autism* **20**, 628–634 (2016).
356. Idring, S. *et al.* Changes in prevalence of autism spectrum disorders in 2001–2011: findings from the Stockholm youth cohort. *J Autism Dev Disord* **45**, 1766–1773 (2015).
357. Yang, J. *et al.* The prevalence of diagnosed tourette syndrome in Canada: A national population-based study. *Mov Disord* **31**, 1658–1663 (2016).
358. Lord, C. *et al.* The Lancet Commission on the future of care and clinical research in autism. *The Lancet* **399**, 271–334 (2022).
359. Bölte, S. *et al.* Sex and gender in neurodevelopmental conditions. *Nature Reviews Neurology* **2023 19:3** **19**, 136–159 (2023).
360. Robinson, E. B., Lichtenstein, P., Anckarsäter, H., Happé, F. & Ronald, A. Examining and interpreting the female protective effect against autistic behavior. *Proc Natl Acad Sci U S A* **110**, 5258–5262 (2013).

361. Dougherty, J. D. *et al.* Can the “female protective effect” liability threshold model explain sex differences in autism spectrum disorder? *Neuron* **110**, 3243–3262 (2022).
362. Van Rijn, S. A review of neurocognitive functioning and risk for psychopathology in sex chromosome trisomy (47,XXY, 47,XXX, 47, XYY). *Curr Opin Psychiatry* **32**, 79 (2019).
363. Slavney, A., Arbiza, L., Clark, A. G. & Keinan, A. Strong Constraint on Human Genes Escaping X-Inactivation Is Modulated by their Expression Level and Breadth in Both Sexes. *Mol Biol Evol* **33**, 384–393 (2016).
364. Brand, B. A., Blesson, A. E. & Smith-Hicks, C. L. The Impact of X-Chromosome Inactivation on Phenotypic Expression of X-Linked Neurodevelopmental Disorders. *Brain Sci* **11**, (2021).
365. Grünblatt, E. *et al.* The involvement of the canonical Wnt-signaling receptor LRP5 and LRP6 gene variants with ADHD and sexual dimorphism: Association study and meta-analysis. *Am J Med Genet B Neuropsychiatr Genet* **180**, 365–376 (2019).
366. Berkel, S. *et al.* Sex Hormones Regulate SHANK Expression. *Front Mol Neurosci* **11**, 391513 (2018).
367. Jung, H. *et al.* Sexually dimorphic behavior, neuronal activity, and gene expression in Chd8-mutant mice. *Nat Neurosci* **21**, 1218–1228 (2018).
368. Quaderi, N. A. *et al.* Opitz G/BBB syndrome, a defect of midline development, is due to mutations in a new RING finger gene on Xp22. *Nat Genet* **17**, 285–291 (1997).
369. Robin, N. H. *et al.* Opitz syndrome is genetically heterogeneous, with one locus on Xp22, and a second locus on 22q11.2. *Nat Genet* **11**, 459–461 (1995).
370. Trockenbacher, A. *et al.* MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that targets phosphatase 2A for degradation. *Nat Genet* **29**, 287–294 (2001).
371. Gaudenz, K. *et al.* Opitz G/BBB syndrome in Xp22: mutations in the MID1 gene cluster in the carboxy-terminal domain. *Am J Hum Genet* **63**, 703–710 (1998).
372. Cox, T. C. *et al.* New mutations in MID1 provide support for loss of function as the cause of X-linked Opitz syndrome. *Hum Mol Genet* **9**, 2553–2562 (2000).
373. Pinson, L. *et al.* Embryonic expression of the human MID1 gene and its mutations in Opitz syndrome. *J Med Genet* **41**, 381–386 (2004).
374. So, J. *et al.* Mild phenotypes in a series of patients with Opitz G/BBB syndrome with MID1 mutations. *Am J Med Genet A* **132A**, 1–7 (2005).
375. Mnayer, L., Khuri, S., Al-Ali Merheby, H., Meroni, G. & Elsas, L. J. A structure-function study of MID1 mutations associated with a mild Opitz phenotype. *Mol Genet Metab* **87**, 198–203 (2006).
376. Ferrentino, R., Bassi, M. T., Chitayat, D., Tabolacci, E. & Meroni, G. MID1 mutation screening in a large cohort of Opitz G/BBB syndrome patients: twenty-nine novel mutations identified. *Hum Mutat* **28**, 206–207 (2007).
377. Palmer, S., Perry, J., Kipling, D. & Ashworth, A. A gene spans the pseudoautosomal boundary in mice. *Proc Natl Acad Sci U S A* **94**, 12030–12035 (1997).
378. Perry, J. *et al.* FXY2/MID2, a gene related to the X-linked Opitz syndrome gene FXY/MID1, maps to Xq22 and encodes a FNIII domain-containing protein that associates with microtubules. *Genomics* **62**, 385–394 (1999).
379. Van Den Veyver, I. B., Cormier, T. A., Jurecic, V., Baldini, A. & Zoghbi, H. Y. Characterization and Physical Mapping in Human and Mouse of a Novel RING Finger Gene in Xp22. *Genomics* **51**, 251–261 (1998).
380. Landry, J.-R. & Mager, D. L. Widely Spaced Alternative Promoters, Conserved between Human and Rodent, Control Expression of the Opitz Syndrome Gene MID1. *Genomics* **80**, 499–508 (2002).
381. Winter, J. *et al.* Regulation of the MID1 protein function is fine-tuned by a complex pattern of alternative splicing. *Hum Genet* **114**, 541–552 (2004).
382. Winter, J. *et al.* Alternative polyadenylation signals and promoters act in concert to control tissue-specific expression of the Opitz Syndrome gene MID1. *BMC Mol Biol* **8**, 105 (2007).

383. Unterbruner, K. *et al.* MicroRNAs miR-19, miR-340, miR-374 and miR-542 regulate MID1 protein expression. *PLoS One* **13**, (2018).
384. Arigoni, M. *et al.* miR-135b Coordinates Progression of ErbB2-Driven Mammary Carcinomas through Suppression of MID1 and MTCH2. *Am J Pathol* **182**, 2058–2070 (2013).
385. Reymond, A. *et al.* The tripartite motif family identifies cell compartments. *EMBO J* **20**, 2140 (2001).
386. Short, K. M. & Cox, T. C. Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. *J Biol Chem* **281**, 8970–8980 (2006).
387. Baldini, R., Mascaro, M. & Meroni, G. The MID1 gene product in physiology and disease. *Gene* **747**, 144655 (2020).
388. Cainarca, S., Messali, S., Ballabio, A. & Meroni, G. Functional characterization of the Opitz syndrome gene product (midin): evidence for homodimerization and association with microtubules throughout the cell cycle. *Hum Mol Genet* **8**, 1387–1396 (1999).
389. Schweiger, S. *et al.* The Opitz syndrome gene product, MID1, associates with microtubules. *Proc Natl Acad Sci U S A* **96**, 2794–2799 (1999).
390. Wright, K. M., Du, H., Dagnachew, M. & Massiah, M. A. Solution structure of the microtubule-targeting COS domain of MID1. *FEBS J* **283**, 3089–3102 (2016).
391. Liu, J., Prickett, T. D., Elliott, E., Meroni, G. & Brautigan, D. L. Phosphorylation and microtubule association of the Opitz syndrome protein mid-1 is regulated by protein phosphatase 2A via binding to the regulatory subunit alpha 4. *Proc Natl Acad Sci U S A* **98**, 6650–6655 (2001).
392. Aranda-Orgillés, B. *et al.* Active transport of the ubiquitin ligase MID1 along the microtubules is regulated by protein phosphatase 2A. *PLoS One* **3**, (2008).
393. Berti, C., Fontanella, B., Ferrentino, R. & Meroni, G. Mig12, a novel Opitz syndrome gene product partner, is expressed in the embryonic ventral midline and co-operates with Mid1 to bundle and stabilize microtubules. *BMC Cell Biol* **5**, (2004).
394. Gholkar, A. A. *et al.* The X-Linked-Intellectual-Disability-Associated Ubiquitin Ligase Mid2 Interacts with Astrin and Regulates Astrin Levels to Promote Cell Division. *Cell Rep* **14**, 180–188 (2016).
395. Zanchetta, M. E., Napolitano, L. M. R., Maddalo, D. & Meroni, G. The E3 ubiquitin ligase MID1/TRIM18 promotes atypical ubiquitination of the BRCA2-associated factor 35, BRAF35. *Biochim Biophys Acta Mol Cell Res* **1864**, 1844–1854 (2017).
396. Zanchetta, M. E. & Meroni, G. Emerging Roles of the TRIM E3 Ubiquitin Ligases MID1 and MID2 in Cytokinesis. *Front Physiol* **10**, (2019).
397. Demir, U., Koehler, A., Schneider, R., Schweiger, S. & Klocker, H. Metformin anti-tumor effect via disruption of the MID1 translational regulator complex and AR downregulation in prostate cancer cells. *BMC Cancer* **14**, 52 (2014).
398. Latta, E. J. & Golding, J. P. Regulation of PP2A activity by Mid1 controls cranial neural crest speed and gangliogenesis. *Mech Dev* **128**, 560–576 (2012).
399. Ciechanover, A. The unravelling of the ubiquitin system. *Nat Rev Mol Cell Biol* **16**, 322–324 (2015).
400. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu Rev Biochem* **67**, 425–479 (1998).
401. Schulman, B. A. & Wade Harper, J. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol* **10**, 319–331 (2009).
402. Kulathu, Y. & Komander, D. Atypical ubiquitylation - the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nat Rev Mol Cell Biol* **13**, 508–523 (2012).
403. Komander, D. & Rape, M. The ubiquitin code. *Annu Rev Biochem* **81**, 203–229 (2012).
404. Huijbregtse, J. M., Scheffner, M., Beaudenon, S. & Howley, P. M. A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc Natl Acad Sci U S A* **92**, 2563 (1995).

405. Joazeiro, C. A. P. & Weissman, A. M. RING finger proteins: mediators of ubiquitin ligase activity. *Cell* **102**, 549–552 (2000).
406. Buetow, L. & Huang, D. T. Structural insights into the catalysis and regulation of E3 ubiquitin ligases. *Nat Rev Mol Cell Biol* **17**, 626–642 (2016).
407. Meroni, G. & Diez-Roux, G. TRIM/RBCC, a novel class of ‘single protein RING finger’ E3 ubiquitin ligases. *Bioessays* **27**, 1147–1157 (2005).
408. Han, X., Du, H. & Massiah, M. A. Detection and characterization of the in vitro e3 ligase activity of the human MID1 protein. *J Mol Biol* **407**, 505–520 (2011).
409. Napolitano, L. M., Jaffray, E. G., Hay, R. T. & Meroni, G. Functional interactions between ubiquitin E2 enzymes and TRIM proteins. *Biochem J* **434**, 309–319 (2011).
410. Di Como, C. J. & Arndt, K. T. Nutrients, via the Tor proteins, stimulate the association of Tap42 with type 2A phosphatases. *Genes Dev* **10**, 1904–1916 (1996).
411. LeNoue-Newton, M. L., Wadzinski, B. E. & Spiller, B. W. The three Type 2A protein phosphatases, PP2Ac, PP4c and PP6c, are differentially regulated by Alpha4. *Biochem Biophys Res Commun* **475**, 64 (2016).
412. McConnell, J. L. *et al.* Alpha4 is a ubiquitin-binding protein that regulates protein serine/threonine phosphatase 2A ubiquitination. *Biochemistry* **49**, 1713–1718 (2010).
413. Watkins, G. R. *et al.* Monoubiquitination promotes calpain cleavage of the protein phosphatase 2A (PP2A) regulatory subunit $\alpha 4$, altering PP2A stability and microtubule-associated protein phosphorylation. *J Biol Chem* **287**, 24207–24215 (2012).
414. Huang, K. & Fingar, D. C. Growing knowledge of the mTOR signaling network. *Semin Cell Dev Biol* **36**, 79–90 (2014).
415. Liu, E., Knutzen, C. A., Krauss, S., Schweiger, S. & Chiang, G. G. Control of mTORC1 signaling by the Opitz syndrome protein MID1. *Proc Natl Acad Sci U S A* **108**, 8680–8685 (2011).
416. Jhanwar-Uniyal, M. *et al.* Diverse signaling mechanisms of mTOR complexes: mTORC1 and mTORC2 in forming a formidable relationship. *Adv Biol Regul* **72**, 51–62 (2019).
417. Hettich, M. M. *et al.* The Anti-Diabetic Drug Metformin Reduces BACE1 Protein Level by Interfering with the MID1 Complex. *PLoS One* **9**, e102420 (2014).
418. Aranda-Orgille, B. *et al.* Protein phosphatase 2A (PP2A)-specific ubiquitin ligase MID1 is a sequence-dependent regulator of translation efficiency controlling 3-phosphoinositide-dependent protein kinase-1 (PDK1). *J Biol Chem* **286**, 39945–39957 (2011).
419. Köhler, A. *et al.* A hormone-dependent feedback-loop controls androgen receptor levels by limiting MID1, a novel translation enhancer and promoter of oncogenic signaling. *Mol Cancer* **13**, (2014).
420. Krauß, S. *et al.* Translation of HTT mRNA with expanded CAG repeats is regulated by the MID1-PP2A protein complex. *Nat Commun* **4**, (2013).
421. Carneiro, B. A., Kaplan, J. B., Altman, J. K., Giles, F. J. & Platanius, L. C. Targeting mTOR signaling pathways and related negative feedback loops for the treatment of acute myeloid leukemia. *Cancer Biol Ther* **16**, 648–656 (2015).
422. Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev* **15**, 3059–3087 (2001).
423. Sasai, N., Toriyama, M. & Kondo, T. Hedgehog Signal and Genetic Disorders. *Front Genet* **10**, (2019).
424. Krauß, S., Foerster, J., Schneider, R. & Schweiger, S. Protein phosphatase 2A and rapamycin regulate the nuclear localization and activity of the transcription factor GLI3. *Cancer Res* **68**, 4658–4665 (2008).
425. Krauß, S. *et al.* Point Mutations in GLI3 Lead to Misregulation of its Subcellular Localization. *PLoS One* **4**, (2009).
426. Schweiger, S. *et al.* The E3 ubiquitin ligase MID1 catalyzes ubiquitination and cleavage of Fu. *J Biol Chem* **289**, 31805–31817 (2014).

427. Pfirrmann, T. *et al.* Hedgehog-dependent E3-ligase Midline1 regulates ubiquitin-mediated proteasomal degradation of Pax6 during visual system development. *Proc Natl Acad Sci U S A* **113**, 10103–10108 (2016).
428. Granata, A. & Quaderi, N. A. The Opitz syndrome gene MID1 is essential for establishing asymmetric gene expression in Hensen's node. *Dev Biol* **258**, 397–405 (2003).
429. Winter, J., Basilicata, M. F., Stemmler, M. P. & Krauss, S. The MID1 protein is a central player during development and in disease. *Front Biosci (Landmark Ed)* **21**, 664–682 (2016).
430. Li, B., Zhou, T. & Zou, Y. Mid1/Mid2 expression in craniofacial development and a literature review of X-linked opitz syndrome. *Mol Genet Genomic Med* **4**, 95–105 (2015).
431. Fontanella, B., Russolillo, G. & Meroni, G. MID1 mutations in patients with X-linked Opitz G/BBB syndrome. *Hum Mutat* **29**, 584–594 (2008).
432. Pearse, R. V. & Young-Pearse, T. L. Lost in translational biology: Understanding sex differences to inform studies of diseases of the nervous system. *Brain Res* **1722**, (2019).
433. Mitra, I. *et al.* Pleiotropic Mechanisms Indicated for Sex Differences in Autism. *PLoS Genet* **12**, (2016).
434. Baron-Cohen, S. *et al.* Why Are Autism Spectrum Conditions More Prevalent in Males? *PLoS Biol* **9**, e1001081 (2011).
435. Schaafsma, S. M. & Pfaff, D. W. Etiologies underlying sex differences in Autism Spectrum Disorders. *Front Neuroendocrinol* **35**, 255–271 (2014).
436. Skuse, D. H. Imprinting, the X-chromosome, and the male brain: Explaining sex differences in the liability to autism. *Pediatric Research* Preprint at <https://doi.org/10.1203/00006450-200001000-00006> (2000).
437. Gockley, J. *et al.* The female protective effect in autism spectrum disorder is not mediated by a single genetic locus. *Mol Autism* **6**, (2015).
438. Gong, X. *et al.* Analysis of X chromosome inactivation in autism spectrum disorders. *Am J Med Genet* **147B**, 830 (2008).
439. Jacquemont, S. *et al.* A Higher Mutational Burden in Females Supports a “Female Protective Model” in Neurodevelopmental Disorders. *Am J Hum Genet* **94**, 415 (2014).
440. Arnold, A. P. A General Theory of Sexual Differentiation. *J Neurosci Res* **95**, 291 (2017).
441. Kelava, I., Chiaradia, I., Pellegrini, L., Kalinka, A. T. & Lancaster, M. A. Androgens increase excitatory neurogenic potential in human brain organoids. *Nature* **602**, 112–116 (2022).
442. Leblond, C. S. *et al.* Operative list of genes associated with autism and neurodevelopmental disorders based on database review. *Mol Cell Neurosci* **113**, (2021).
443. Lyon, M. F. Gene Action in the X-chromosome of the Mouse (*Mus musculus* L.). *Nature* **196**:4773 **190**, 372–373 (1961).
444. Żylicz, J. J. *et al.* The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell* **176**, 182–197.e23 (2019).
445. Jégu, T., Aeby, E. & Lee, J. T. The X chromosome in space. *Nature Reviews Genetics* Preprint at <https://doi.org/10.1038/nrg.2017.17> (2017).
446. Galupa, R. & Heard, E. X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. *Annu Rev Genet* **52**, 535–566 (2018).
447. Loda, A., Collombet, S. & Heard, E. Gene regulation in time and space during X-chromosome inactivation. *Nat Rev Mol Cell Biol* **23**, 231–249 (2022).
448. Berletch, J. B. *et al.* Escape from X Inactivation Varies in Mouse Tissues. *PLoS Genet* (2015) doi:10.1371/journal.pgen.1005079.

449. Wu, H. *et al.* Cellular Resolution Maps of X Chromosome Inactivation: Implications for Neural Development, Function, and Disease. *Neuron* (2014) doi:10.1016/j.neuron.2013.10.051.
450. Vangipuram, M., Ting, D., Kim, S., Diaz, R. & Schüle, B. Skin punch biopsy explant culture for derivation of primary human fibroblasts. *J Vis Exp* (2013) doi:10.3791/3779.
451. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823 (2013).
452. Yumlu, S., Bashir, S., Stumm, J. & Kühn, R. Efficient Gene Editing of Human Induced Pluripotent Stem Cells Using CRISPR/Cas9. *Methods Mol Biol* **1961**, 137–151 (2019).
453. Taylor, S. C. *et al.* The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time. *Trends Biotechnol* **37**, 761–774 (2019).
454. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10–12 (2011).
455. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
456. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).
457. Poplin, R. *et al.* A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* **36**, 983 (2018).
458. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 1–14 (2016).
459. Aguet, F. *et al.* The impact of sex on gene expression across human tissues. *Science* (1979) **369**, (2020).
460. Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).
461. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**:7539 **518**, 317–330 (2015).
462. Yen, A. & Kellis, M. Systematic chromatin state comparison of epigenomes associated with diverse properties including sex and tissue type. *Nature Communications* **2015** **6**:1 **6**, 1–13 (2015).
463. Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nature* (2016) doi:10.1038/nature18589.
464. Hauth, A. *et al.* Escape from X inactivation is directly modulated by levels of Xist non-coding RNA. *bioRxiv* 2024.02.22.581559 (2024) doi:10.1101/2024.02.22.581559.
465. Carrel, L. & Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400–404 (2005).
466. Vallot, C. *et al.* Erosion of X chromosome inactivation in human pluripotent cells initiates with XACT coating and depends on a specific heterochromatin landscape. *Cell Stem Cell* **16**, 533–546 (2015).
467. Mekhoubad, S. *et al.* Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell* **10**, 595–609 (2012).
468. Bar, S., Seaton, L. R., Weissbein, U., Eldar-Geva, T. & Benvenisty, N. Global Characterization of X Chromosome Inactivation in Human Pluripotent Stem Cells. *CellReports* **27**, 20–29.e3 (2019).
469. Patel, S. *et al.* Human Embryonic Stem Cells Do Not Change Their X Inactivation Status during Differentiation. *Cell Rep* **18**, 54–67 (2017).
470. Raposo, A. C. *et al.* Erosion of X-Chromosome Inactivation in female hiPSCs is heterogeneous and persists during differentiation. *bioRxiv* 2024.03.15.585169 (2024) doi:10.1101/2024.03.15.585169.
471. Fukuda, A. *et al.* De novo DNA methyltransferases DNMT3A and DNMT3B are essential for XIST silencing for erosion of dosage compensation in pluripotent stem cells. *Stem Cell Reports* **16**, 2138 (2021).
472. Raposo, A. C., Casanova, M., Gendrel, A. V. & da Rocha, S. T. The tandem repeat modules of Xist lncRNA: a swiss army knife for the control of X-chromosome inactivation. *Biochem Soc Trans* **49**, 2549–2560 (2021).

473. Geens, M. *et al.* Female human pluripotent stem cells rapidly lose X chromosome inactivation marks and progress to a skewed methylation pattern during culture. *Mol Hum Reprod* **22**, 285–298 (2016).
474. Motosugi, N. *et al.* Deletion of lncRNA XACT does not change expression dosage of X-linked genes, but affects differentiation potential in hPSCs. *Cell Rep* **35**, (2021).
475. Whitworth, D. J. & Pask, A. J. The X factor: X chromosome dosage compensation in the evolutionarily divergent monotremes and marsupials. *Semin Cell Dev Biol* **56**, 117–121 (2016).
476. Mahadevaiah, S. K., Sangrithi, M. N., Hirota, T. & Turner, J. M. A. A single-cell transcriptome atlas of marsupial embryogenesis and X inactivation. *Nature* **586**, 612–617 (2020).
477. Grant, J. *et al.* Rxs is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature* **487**, 254–258 (2012).
478. Okamoto, I. *et al.* Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472**, 370–374 (2011).
479. DeCasien, A. R., Guma, E., Liu, S. & Raznahan, A. Sex differences in the human brain: a roadmap for more careful analysis and interpretation of a biological reality. *Biol Sex Differ* **13**, (2022).
480. Pottmeier, P., Nikolantonaki, D., Lanner, F., Peuckert, C. & Jazin, E. Sex-biased gene expression during neural differentiation of human embryonic stem cells. *Front Cell Dev Biol* **12**, 1341373 (2024).
481. Wingo, A. P. *et al.* Sex differences in brain protein expression and disease. *Nature Medicine* **2023** *29*:9 **29**, 2224–2232 (2023).
482. Lahita, R. G. Sex and gender influence on immunity and autoimmunity. *Front Immunol* **14**, (2023).
483. Calabrò, A., Accardi, G., Aiello, A., Caruso, C. & Candore, G. Sex and gender affect immune aging. *Frontiers in Aging* **4**, (2023).
484. Balaton, B. P. & Brown, C. J. Escape Artists of the X Chromosome. *Trends Genet* **32**, 348–359 (2016).
485. Carrel, L. & Brown, C. J. When the Lyon(ized chromosome) roars: ongoing expression from an inactive X chromosome. *Philos Trans R Soc Lond B Biol Sci* **372**, (2017).
486. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).
487. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
488. Steeg, R. *et al.* EBiSC best practice: How to ensure optimal generation, qualification, and distribution of iPSC lines. *Stem Cell Reports* **16**, 1853–1867 (2021).
489. Ben-David, U. & Benvenisty, N. The tumorigenicity of human embryonic and induced pluripotent stem cells. *Nat Rev Cancer* **11**, 268–277 (2011).
490. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc Jpn Acad Ser B Phys Biol Sci* **85**, 348–362 (2009).
491. Doss, M. X. & Sachinidis, A. Current Challenges of iPSC-Based Disease Modeling and Therapeutic Implications. *Cells* **8**, (2019).
492. Roth, J. G. *et al.* 16p11.2 microdeletion imparts transcriptional alterations in human iPSC-derived models of early neural development. *Elife* **9**, 1–34 (2020).
493. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
494. Park, I. H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).

495. Xing, Q. R. *et al.* Diversification of reprogramming trajectories revealed by parallel single-cell transcriptome and chromatin accessibility sequencing. *Sci Adv* **6**, (2020).
496. Vallier, L., Alexander, M. & Pedersen, R. A. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J Cell Sci* **118**, 4495–4509 (2005).
497. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
498. Brons, I. G. M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195 (2007).
499. Adewumi, O. *et al.* Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nature Biotechnology* **25**, 803–816 (2007).
500. Hatano, S. Y. *et al.* Pluripotential competence of cells associated with Nanog activity. *Mech Dev* **122**, 67–79 (2005).
501. Suresh Babu, S. *et al.* Characterization of human induced pluripotent stem cells: Current approaches, challenges, and future solutions. *Biotechnol Rep (Amst)* **37**, (2023).
502. Avior, Y., Biancotti, J. C. & Benvenisty, N. TeratoScore: Assessing the Differentiation Potential of Human Pluripotent Stem Cells by Quantitative Expression Analysis of Teratomas. *Stem Cell Reports* **4**, 967–974 (2015).
503. Müller, F. J. *et al.* A bioinformatic assay for pluripotency in human cells. *Nat Methods* **8**, 315 (2011).
504. Tsankov, A. M. *et al.* A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells. *Nat Biotechnol* **33**, 1182–1192 (2015).
505. Sánchez-González, J. L. *et al.* Analysis of Telomere Length and Its Implication in Neurocognitive Functions in Elderly Women. *J Clin Med* **11**, (2022).
506. Zhang, S. *et al.* Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. *Science* **369**, 561–565 (2020).
507. Qiu, J. J., Liu, Y. N., Wei, H., Zeng, F. & Yan, J. Bin. Single-cell RNA sequencing of neural stem cells derived from human trisomic iPSCs reveals the abnormalities during neural differentiation of Down syndrome. *Front Mol Neurosci* **16**, (2023).
508. Chambers, S. M. *et al.* Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnology* **27**, 275–280 (2009).
509. Qi, Y. *et al.* Combined small-molecule inhibition accelerates the derivation of functional cortical neurons from human pluripotent stem cells. *Nat Biotechnol* **35**, 154–163 (2017).
510. Morshead, C. M. *et al.* Neural stem cells in the adult mammalian forebrain: a relatively quiescent subpopulation of subependymal cells. *Neuron* **13**, 1071–1082 (1994).
511. Reynolds, B. A. & Weiss, S. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science* **255**, 1707–1710 (1992).
512. Dahlstrand, J., Lardelli, M. & Lendahl, U. Nestin mRNA expression correlates with the central nervous system progenitor cell state in many, but not all, regions of developing central nervous system. *Developmental Brain Research* **84**, 109–129 (1995).
513. Frederikson, K. & McKay, R. D. G. Proliferation and differentiation of rat neuroepithelial precursor cells in vivo. *J Neurosci* **8**, 1144–1151 (1988).
514. Lendahl, U., Zimmerman, L. B. & McKay, R. D. G. CNS stem cells express a new class of intermediate filament protein. *Cell* **60**, 585–595 (1990).
515. Kikkawa, T. *et al.* The role of Pax6 in brain development and its impact on pathogenesis of autism spectrum disorder. *Brain Res* **1705**, 95–103 (2019).
516. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev* **17**, 126–140 (2003).

517. Thomson, M. *et al.* Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145**, 875–889 (2011).
518. Pevny, L. H. & Nicolis, S. K. Sox2 roles in neural stem cells. *Int J Biochem Cell Biol* **42**, 421–424 (2010).
519. Kim, J. B. *et al.* Oct4-induced pluripotency in adult neural stem cells. *Cell* **136**, 411–419 (2009).
520. Brewer, G. J., Torricelli, J. R., Evege, E. K. & Price, P. J. Optimized survival of hippocampal neurons in B27-supplemented Neurobasal, a new serum-free medium combination. *J Neurosci Res* **35**, 567–576 (1993).
521. Tan, B. T. *et al.* Retinoic acid induced the differentiation of neural stem cells from embryonic spinal cord into functional neurons in vitro. *Int J Clin Exp Pathol* **8**, 8129 (2015).
522. Janesick, A., Wu, S. C. & Blumberg, B. Retinoic acid signaling and neuronal differentiation. *Cell Mol Life Sci* **72**, 1559–1576 (2015).
523. Del Corral, R. D. *et al.* Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron* **40**, 65–79 (2003).
524. Kwon, S. E. & Chapman, E. R. Synaptophysin regulates the kinetics of synaptic vesicle endocytosis in central neurons. *Neuron* **70**, 847 (2011).
525. Menezes, J. R. L. & Luskin, M. B. Expression of neuron-specific tubulin defines a novel population in the proliferative layers of the developing telencephalon. *J Neurosci* **14**, 5399–5416 (1994).
526. Binder, L. I., Frankfurter, A. & Rebhun, L. I. The distribution of tau in the mammalian central nervous system. *J Cell Biol* **101**, 1371–1378 (1985).
527. Silva, S. S., Rowntree, R. K., Mekhoubad, S. & Lee, J. T. X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proc Natl Acad Sci U S A* **105**, 4820–4825 (2008).
528. Yamanaka, S. Pluripotent Stem Cell-Based Cell Therapy—Promise and Challenges. *Cell Stem Cell* **27**, 523–531 (2020).
529. Sauteraud, R. *et al.* Inferring genes that escape X-Chromosome inactivation reveals important contribution of variable escape genes to sex-biased diseases. *Genome Res* **31**, 1629–1637 (2021).
530. Zito, A. *et al.* Escape from X-inactivation in twins exhibits intra- and inter-individual variability across tissues and is heritable. *PLoS Genet* **19**, e1010556 (2023).
531. Cotton, A. M. *et al.* Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol* (2013) doi:10.1186/gb-2013-14-11-r122.
532. Amos-Landgraf, J. M. *et al.* X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. *Am J Hum Genet* **79**, 493–499 (2006).
533. Cloutier, M. *et al.* Preventing erosion of X-chromosome inactivation in human embryonic stem cells. *Nature Communications* 2022 13:1 **13**, 1–18 (2022).
534. Jakobsson, E. *et al.* Towards a Unified Understanding of Lithium Action in Basic Biology and its Significance for Applied Biology. *J Membr Biol* **250**, 587–604 (2017).
535. Bansal, P., Ahern, D. T., Kondaveeti, Y., Qiu, C. W. & Pinter, S. F. Contiguous erosion of the inactive X in human pluripotency concludes with global DNA hypomethylation. *Cell Rep* **35**, (2021).
536. Yokobayashi, S. *et al.* Inherent genomic properties underlie the epigenomic heterogeneity of human induced pluripotent stem cells. *Cell Rep* **37**, (2021).
537. Vallot, C. *et al.* XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. *Cell Stem Cell* **20**, 102–111 (2017).
538. Motosugi, N. *et al.* De-erosion of X chromosome dosage compensation by the editing of XIST regulatory regions restores the differentiation potential in hPSCs. *Cell Reports Methods* **2**, (2022).
539. de Sousa, J. A. *et al.* Epigenetic dynamics during capacitation of naïve human pluripotent stem cells. *Sci Adv* **9**, (2023).

540. Cali, C. P., Park, D. S. & Lee, E. B. Targeted DNA methylation of neurodegenerative disease genes via homology directed repair. *Nucleic Acids Res* **47**, 11609–11622 (2019).
541. Brenes, A. J. *et al.* Erosion of human X chromosome inactivation causes major remodeling of the iPSC proteome. *Cell Rep* **35**, (2021).
542. Balaton, B. P., Fornes, O., Wasserman, W. W. & Brown, C. J. Cross-species examination of X-chromosome inactivation highlights domains of escape from silencing. *Epigenetics Chromatin* **14**, 1–17 (2021).
543. Willard, H. F., Brown, C. J., Carrel, L., Hendrich, B. & Miller, A. P. Epigenetic and Chromosomal Control of Gene Expression: Molecular and Genetic Analysis of X Chromosome Inactivation. *Cold Spring Harb Symp Quant Biol* **58**, 315–322 (1993).
544. Clemson, C. M., Chow, J. C., Brown, C. J. & Lawrence, J. B. Stabilization and Localization of Xist RNA are Controlled by Separate Mechanisms and are Not Sufficient for X Inactivation. *J Cell Biol* **142**, 13 (1998).
545. San Roman, A. K. *et al.* The human inactive X chromosome modulates expression of the active X chromosome. *Cell genomics* **3**, (2023).
546. Roman, A. K. S. *et al.* The human Y and inactive X chromosomes similarly modulate autosomal gene expression. *bioRxiv* (2023) doi:10.1101/2023.06.05.543763.
547. Balaton, B. P. & Brown, C. J. Contribution of genetic and epigenetic changes to escape from X-chromosome inactivation. *Epigenetics Chromatin* **14**, (2021).
548. Boudewijns, M., van Dongen, J. J. M. & Langerak, A. W. The Human Androgen Receptor X-Chromosome Inactivation Assay for Clonality Diagnostics of Natural Killer Cell Proliferations. *J Mol Diagn* **9**, 337 (2007).
549. Peaherrera, M. S. *et al.* Patterns of placental development evaluated by X chromosome inactivation profiling provide a basis to evaluate the origin of epigenetic variation. *Hum Reprod* **27**, 1745–1753 (2012).
550. Werner, J. M., Ballouz, S., Hover, J. & Gillis, J. Variability of cross-tissue X-chromosome inactivation characterizes timing of human embryonic lineage specification events. *Dev Cell* **57**, 1995-2008.e5 (2022).
551. Hatakeyama, C. *et al.* The dynamics of X-inactivation skewing as women age. *Clin Genet* **66**, 327–332 (2004).
552. Plenge, R. M., Stevenson, R. A., Lubs, H. A., Schwartz, C. E. & Willard, H. F. Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. *Am J Hum Genet* **71**, 168–173 (2002).
553. Nesterova, T. B. *et al.* Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nat Commun* **10**, (2019).
554. Wainer Katsir, K. & Linial, M. Human genes escaping X-inactivation revealed by single cell expression data. *BMC Genomics* **20**, 1–17 (2019).
555. Basta, M. & Pandya, A. M. Genetics, X-Linked Inheritance. *StatPearls* (2023).
556. Alfeghaly, C. *et al.* XIST dampens X chromosome activity in a SPEN-dependent manner during early human development. *Nature Structural & Molecular Biology* **2024** 1–12 (2024) doi:10.1038/s41594-024-01325-3.
557. Bellott, D. W. *et al.* Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014).
558. Pacini, G. *et al.* Integrated analysis of Xist upregulation and X-chromosome inactivation with single-cell and single-allele resolution. *Nature Communications* **2021** 12:1 **12**, 1–17 (2021).
559. Simon, M. D. *et al.* High-resolution Xist binding maps reveal 2-step spreading during X-inactivation. *Nature* **504**, 465 (2013).
560. Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, (2013).
561. Murakami, K. *et al.* Identification of the chromatin regions coated by non-coding Xist RNA. *Cytogenet Genome Res* **125**, 19–25 (2009).

562. McCarthy, M. M. Sex differences in the developing brain as a source of inherent risk. *Dialogues Clin Neurosci* **18**, 361 (2016).
563. Wen, X. *et al.* Single-cell multiplex chromatin and RNA interactions in ageing human brain. *Nature* **2024 628:8008 628**, 648–656 (2024).
564. Bergero, R. & Charlesworth, D. The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol* **24**, 94–102 (2009).
565. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
566. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **2003 423:6942 423**, 825–837 (2003).
567. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
568. Rice, W. R. SEX CHROMOSOMES AND THE EVOLUTION OF SEXUAL DIMORPHISM. *Evolution (N Y)* **38**, 735–742 (1984).
569. Lemaitre, C. *et al.* Footprints of Inversions at Present and Past Pseudoautosomal Boundaries in Human Sex Chromosomes. *Genome Biol Evol* **1**, 56–66 (2009).
570. Ning, Z. *et al.* Regulation of SPRY3 by X chromosome and PAR2-linked promoters in an autism susceptibility region. *Hum Mol Genet* **24**, 5126–5141 (2015).
571. Balaton, B. P., Cotton, A. M. & Brown, C. J. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biol Sex Differ* (2015) doi:10.1186/s13293-015-0053-7.
572. Posynick, B. J. & Brown, C. J. Escape From X-Chromosome Inactivation: An Evolutionary Perspective. *Front Cell Dev Biol* **7**, (2019).
573. Sanchez-Roige, S. *et al.* A mutant allele of glycoprotein M6-B (Gpm6b) facilitates behavioral flexibility but increases delay discounting. *Genes Brain Behav* **21**, (2022).
574. Zhang, X. *et al.* Glycoprotein M6B Interacts with TβRI to Activate TGF-β-Smad2/3 Signaling and Promote Smooth Muscle Cell Differentiation. *Stem Cells* **37**, 190–201 (2019).
575. Choi, K. M., Kim, J. Y. & Kim, Y. Distribution of the Immunoreactivity for Glycoprotein M6B in the Neurogenic Niche and Reactive Glia in the Injury Penumbra Following Traumatic Brain Injury in Mice. *Exp Neurobiol* **22**, 277 (2013).
576. Mita, S. *et al.* Transcallosal Projections Require Glycoprotein M6-Dependent Neurite Growth and Guidance. *Cerebral Cortex* **25**, 4111–4125 (2015).
577. Fuchsova, B., Alvarez Juliá, A., Rizavi, H. S., Frasch, A. C. & Pandey, G. N. Altered expression of neuroplasticity-related genes in the brain of depressed suicides. *Neuroscience* **299**, 1–17 (2015).
578. Berletch, J. B., Yang, F. & Disteché, C. M. Escape from X inactivation in mice and humans. *Genome Biology* vol. 11 213 Preprint at <https://doi.org/10.1186/gb-2010-11-6-213> (2010).
579. Wu, H. *et al.* Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease. *Neuron* **81**, 103–119 (2014).
580. Short, K. M., Hopwood, B., Yi, Z. & Cox, T. C. MID1 and MID2 homo- and heterodimerise to tether the rapamycin-sensitive PP2A regulatory subunit, alpha 4, to microtubules: implications for the clinical variability of X-linked Opitz GBBB syndrome and other developmental disorders. *BMC Cell Biol* **3**, (2002).

8 Curriculum vitae

9 Acknowledgements

