

**Autora:**

Sarah Robert

**Director:**

Dr. Álvaro Rada-Iglesias

## Tesis Doctoral

**MODELIZACIÓN *IN VITRO* DE LA  
HOLOPROSENCEFALIA RELACIONADA CON ZIC2  
EN HUMANOS REVELA GENES DIANA SENSIBLES  
A LA DOSIS DE ZIC2**

## PhD Thesis

***IN VITRO* DISEASE MODELLING OF HUMAN ZIC2-  
RELATED HOLOPROSENCEPHALY UNCOVERS  
DOSAGE-SENSITIVE ZIC2 TARGET GENES**



Escuela de **Doctorado**

La realización de este proyecto de tesis doctoral ha sido posible gracias a la obtención de financiación proveniente de las siguientes fuentes:

- Programa de Jóvenes Investigadores de la EMBO (EMBO YIP).
- Planes estatales PGC2018- 095301-B-I00 y PID2021-123030NB-I00 financiados por MCIN/AEI/10.13039/501 100 011 033 y por 'ERDF A way of making Europe'.
- RED2022-134100-T (REDEVNEURAL 3.0) financiada por MCIN/AEI/10.13039/501100011033.
- Consejo Europeo de Investigación, ERC CoG 'PoisedLogic' [862 022].
- ENHPATHY' H2020-MSCA-ITN-2019-860002 financiada por la Comisión Europea.



## Acknowledgement

First of all, I would like to thank Eloisa Herrera, Robin Andersson and Ignacio Varela, who kindly accepted to be members of my tribunal. I am truly grateful for the time spent reading the manuscript and look forward to an enriching discussion. I would also like to thank Marian Ros and Ana Rosa Palanca for the time dedicated to the annual evaluation and their suggestions during this PhD journey.

Álvaro, I will never thank you enough for giving me the opportunity to do this PhD in your team, for your kindness and supervision. I could not wish for a better environment to do my PhD.

I was also fortunate to be part of the Enhpathy consortium and meet so many talented people. Thank you all for the inspiration, support, and great memories.

I would not have been able to complete this PhD without the unconditional support from the lab members, who have become true friends.

Maria, Victor, you were the first people I got to meet when I joined the lab. Thank you for showing me Cantabria, for the funny stories and for all the support in and outside the lab. Marianna, thank you for always being present, in the good and bad moments. Thais, thank you for your unlimited kindness and for never hesitating to help out. Lara, you know how to mess with people in the best and positive way. Thank you for always bringing joy to the lab. Mitsuru, you are truly inspiring, thank you for joining the lab all the way from Japan. Lucio, thank you for always answering my questions and for your kindness. Patri, thank you for your help and for always caring. Helena, thank you for joining the lab, for your friendship and for the support. Endika, thank you for your critical thinking and always insightful comments. Dani, your visits to the lab always bring light and smiles. I love you all.

Javi, gracias por apoyarme siempre.

Ces remerciements ne seraient pas complets sans mentionner ma famille. Maman, Thomas, Yasmine, Hamid, Dounia, Samira, merci pour tout.

# Table of contents

|                                                                             |    |
|-----------------------------------------------------------------------------|----|
| Acknowledgement .....                                                       | 4  |
| Abstract.....                                                               | 7  |
| Resumen .....                                                               | 8  |
| Abbreviations.....                                                          | 9  |
| List of figures .....                                                       | 11 |
| List of tables .....                                                        | 12 |
| 1. Introduction .....                                                       | 13 |
| 1.1. Transcription factors are master regulators of gene expression .....   | 14 |
| 1.2. Transcription factors and disease .....                                | 16 |
| 1.3. Early brain development.....                                           | 17 |
| 1.4. Early brain patterning .....                                           | 18 |
| 1.5. ZIC family of transcription factors .....                              | 23 |
| 1.6. ZIC2-related HPE .....                                                 | 25 |
| 1.7. ZIC2 as an enhancer binding factor.....                                | 26 |
| 1.8. Modeling early brain development <i>in vitro</i> .....                 | 27 |
| 1.9. Genomics to study ZIC2 function.....                                   | 28 |
| 2. Objectives.....                                                          | 29 |
| 3. Materials and Methods .....                                              | 30 |
| 3.1. Material .....                                                         | 31 |
| 3.1.1. Equipment.....                                                       | 31 |
| 3.1.2. Cell lines .....                                                     | 32 |
| 3.1.3. Cell culture reagents and media .....                                | 32 |
| 3.1.4. Molecular biology kits, reagents, enzymes and antibodies .....       | 34 |
| 3.1.5. Molecular cloning .....                                              | 36 |
| 3.2. Methods .....                                                          | 37 |
| 3.2.1. Culture of hiPSC .....                                               | 37 |
| 3.2.2. Differentiation of hiPSC into anterior neural progenitor cells ..... | 37 |
| 3.2.3. Generation of transgenic hiPSC lines with CRISPR-Cas9 .....          | 38 |
| a. gRNA design .....                                                        | 38 |
| b. Generation of the gRNA-CRISPR-Cas9 vectors .....                         | 40 |
| c. Transfection of hiPSC and clone isolation .....                          | 41 |
| 3.2.4. Molecular biology methods .....                                      | 42 |
| a. Protein extraction and western blot.....                                 | 44 |
| b. RNA extraction, cDNA synthesis and RNA-seq samples.....                  | 44 |

|            |                                                                                                                           |     |
|------------|---------------------------------------------------------------------------------------------------------------------------|-----|
| c.         | Single-cell RNA-seq .....                                                                                                 | 45  |
| d.         | Chromatin immunoprecipitation and ChIP-seq .....                                                                          | 45  |
| e.         | Polymerase Chain Reaction.....                                                                                            | 46  |
| 3.2.5.     | Data analysis .....                                                                                                       | 51  |
| a.         | qPCR analysis.....                                                                                                        | 52  |
| b.         | RNA-seq analysis.....                                                                                                     | 52  |
| c.         | ChIP-seq analysis .....                                                                                                   | 53  |
| d.         | Single-cell RNA-seq analysis .....                                                                                        | 54  |
| e.         | Integration of ChIP-seq and RNA-seq data .....                                                                            | 54  |
| 4.         | Results .....                                                                                                             | 55  |
| 4.1.       | Differentiation of WT hiPSC into anterior neural progenitor cells (AntNPC) .....                                          | 56  |
| 4.2.       | Genetic engineering of hiPSC .....                                                                                        | 60  |
| 4.3.       | Characterization of molecular defects in <i>ZIC2</i> <sup>-/-</sup> cells.....                                            | 66  |
| 4.4.       | <i>ZIC</i> genes are strongly up-regulated in <i>ZIC2</i> <sup>-/-</sup> neural progenitors compared to WT<br>72          |     |
| 4.5.       | Characterization of <i>ZIC2</i> binding profiles during AntNPC differentiation .....                                      | 74  |
| 4.6.       | Characterization of <i>ZIC2</i> bound regions gaining H3K27ac in <i>ZIC2</i> <sup>-/-</sup> cells.....                    | 77  |
| 4.7.       | Characterization of <i>ZIC2</i> bound regions losing H3K27ac in <i>ZIC2</i> <sup>-/-</sup> cells.....                     | 81  |
| 4.8.       | The positional identities of neural progenitors are shifted upon loss of <i>ZIC2</i> ....                                 | 85  |
| 4.9.       | Investigating how the loss of <i>ZIC2</i> affects the responsiveness to hedgehog<br>signaling in neural progenitors ..... | 87  |
| 4.10.      | Reduced <i>ZIC2</i> dosage also leads to significant transcriptional defects in AntNPCs<br>92                             |     |
| 4.11.      | Future experiments .....                                                                                                  | 94  |
| 5.         | Discussion .....                                                                                                          | 97  |
| 5.1.       | Modeling human forebrain development <i>in vitro</i> .....                                                                | 98  |
| 5.2.       | <i>ZIC2</i> in hiPSC .....                                                                                                | 99  |
| 5.3.       | Dorsal-ventral patterning of forebrain-like is impaired in <i>ZIC2</i> <sup>-/-</sup> neural progenitors<br>99            |     |
| 5.4.       | <i>ZIC2</i> as an enhancer binding factor .....                                                                           | 102 |
| 5.5.       | Human and mouse show different sensitivity to <i>ZIC2</i> dosage during forebrain<br>development .....                    | 103 |
| 6.         | Conclusions.....                                                                                                          | 104 |
| References | .....                                                                                                                     | 107 |

## Abstract

During development, transcription factors (TFs) play a key role in ensuring the robust establishment of gene expression programs and, thus, of cellular identities. In this work, we used a human induced pluripotent stem cells (hiPSCs) *in vitro* differentiation system that models forebrain development to investigate the role of the *ZIC2* gene, a member of the ZIC family of TFs, during the induction of anterior neural identities. The relevance of the *ZIC2* gene is highlighted by its association with holoprosencephaly (HPE), a forebrain congenital disorder affecting the proper separation of the cerebral hemispheres. Despite the medical and developmental relevance of *ZIC2*, little is known about the regulatory network controlled by this TF during forebrain development, particularly in humans. We therefore took advantage of hiPSCs and differentiated them into anterior neural progenitor cells (AntNPCs) with a forebrain-like identity to study the regulatory network controlled by *ZIC2* upon neural induction. Combining CRISPR-Cas9 genome editing approaches and genomic methods (RNA-seq, ChIP-seq and scRNA-seq), we showed that (i) our differentiation system of hiPSCs into AntNPCs recapitulates important aspects of dorsal-ventral and rostro-caudal forebrain patterning; (ii) in the absence of *ZIC2*, AntNPCs showed increased expression of rostral telencephalon markers (e.g. *SIX3*, *FOXG1*, *FGF8*) and reduced levels of WNT signaling and cortical hem markers (e.g. *WNT2B*, *WNT3A*, *LMX1A*), (iii) the loss of *ZIC2* sensitized AntNPC to SHH signaling; (iv) *ZIC2* binds to and is required for the activation of distal enhancers associated with major forebrain patterning regulators; (v) AntNPC display *ZIC2* haploinsufficiency, as the loss of one *ZIC2* allele already resulted in severe expression defect for major forebrain patterning genes. Altogether, this work provides novel insights into the regulatory network controlled by *ZIC2* during human forebrain patterning, which in turn can help understanding the molecular basis of HPE.

## Resumen

Durante el desarrollo, los factores de transcripción (TFs) juegan un papel clave en garantizar el establecimiento robusto de los programas de expresión génica y, por lo tanto, de las identidades celulares. En este trabajo, utilizamos un sistema de diferenciación in vitro de células madre pluripotentes inducidas humanas (hiPSCs) que modela el desarrollo del prosencéfalo para investigar el papel del gen *ZIC2*, un miembro de la familia de TFs ZIC, durante la inducción de identidades neuronales anteriores. La relevancia del gen *ZIC2* se destaca por su asociación con la holoprosencefalia (HPE), un trastorno congénito del prosencéfalo que afecta la correcta separación de los hemisferios cerebrales. A pesar de la importancia médica y en el desarrollo de *ZIC2*, se sabe poco sobre la red reguladora controlada por este TF durante el desarrollo del prosencéfalo, particularmente en humanos. Por ello, aprovechamos el sistema de diferenciación de hiPSCs hacia células progenitoras neuronales anteriores (AntNPCs) con identidad similar a la del prosencéfalo para estudiar la red reguladora controlada por *ZIC2* durante la inducción neural. Combinando métodos de edición genética con CRISPR-Cas9 y análisis genómicos (RNA-seq, ChIP-seq y scRNA-seq), mostramos que (i) nuestro sistema de diferenciación de hiPSCs a AntNPCs recapitula aspectos clave del patrón dorsal-ventral y rostro-caudal del prosencéfalo; (ii) en ausencia de *ZIC2*, las AntNPCs presentan una mayor expresión de marcadores del telencéfalo rostral (*SIX3*, *FOXG1*, *FGF8*) y niveles reducidos de señalización WNT y de marcadores del cortical hem (*WNT2B*, *WNT3A*, *LMX1A*); (iii) la pérdida de *ZIC2* sensibiliza a las AntNPCs a la señalización de SHH; (iv) *ZIC2* se une y es necesario para la activación de potenciadores distales asociados con reguladores clave del patrón del prosencéfalo; (v) las AntNPCs exhiben haploinsuficiencia de *ZIC2*, ya que la pérdida de un alelo de *ZIC2* provoca defectos severos en la expresión de genes clave para el patrón del prosencéfalo. En conjunto, este trabajo proporciona nuevos conocimientos sobre la red reguladora controlada por *ZIC2* durante la organización del prosencéfalo humano, lo que a su vez puede ayudar a comprender la base molecular de la HPE.

## Abbreviations

AmpR: Ampicillin Resistance

ANR: Anterior Neural Ridge

AntNPC: Anterior Neural Progenitor Cell

BCA: Bicinchoninic Acid

bp: base pairs

BSA: Bovine Serum Albumin

Cas9: CRISPR-associated protein 9

cDNA: Complementary DNA

CGI: CpG Island

ChIP: Chromatin Immunoprecipitation

COF: Co-Factor

CPM: Counts Per Million

CRE: Cis-Regulatory Element

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

Ct: Cycle threshold

DEG: Differentially Expressed Gene

DEL: Deletion

DMSO: Dimethyl Sulfoxide

DNA: Deoxyribonucleic Acid

E. coli: Escherichia coli

eRNA: enhancer RNA

FBS: Fetal Bovine Serum

FDR: False Discovery Rate

FHA: Flag-HA

FPKM: Fragments Per Kilobase of transcript per Million mapped reads

gDNA: genomic DNA

GO: Gene Ontology

GREAT: Genomic Regions Enrichment of Annotations Tool

gRNA: guide RNA

hESC: Human Embryonic Stem Cell

hiPSC: Human Induced Pluripotent Stem Cell

HPE: Holoprosencephaly

HRP: Horseradish Peroxidase

INV: Inversion

kb: kilobase

KSR: Knockout Serum Replacement

LB1/2/3: Lysis Buffer 1, 2, and 3

LDN: LDN193189 (BMP signaling inhibitor)

MEM NEAA: Minimum Essential Medium Non-Essential Amino Acids

mESC: Mouse Embryonic Stem Cell

MIHV: Middle Interhemispheric Variant

NHEJ: Non-Homologous End Joining

ON: overnight

Opa: Odd-paired

p-adj: adjusted p-value

PAGE: Polyacrylamide Gel Electrophoresis

PBS: Phosphate-Buffered Saline

PCA: Principal Component Analysis

PCR: Polymerase Chain Reaction

PuroR: Puromycin Resistance

qPCR: quantitative Polymerase Chain Reaction

RNA: Ribonucleic Acid

RT: Room temperature

RT-qPCR: Reverse Transcription Quantitative Polymerase Chain Reaction

SAG: Smoothed Agonist (Hedgehog signaling activator)

SB: SB431542 (TGF $\beta$  signaling inhibitor)

scRNA-seq: Single-Cell RNA Sequencing

SDS: Sodium Dodecyl Sulfate

SNP: Single Nucleotide Polymorphism

TF: Transcription Factor

TFBS: Transcription Factor Binding Site

TSS: Transcription Start Site

UMAP: Uniform Manifold Approximation  
and Projection

UMI: Unique Molecular Identifier

UTR: Untranslated Region

vs: versus

WT: Wild-type

XAV: XAV939, WNT pathway inhibitor

ZIC: Zinc Finger of the Cerebellum

ZLI: Zona Limitans Intrathalamica

## List of figures

|                                                                                                                                                                                            |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1.1: Schematic representation of the main types of cis-regulatory elements. ....                                                                                                    | 15 |
| Figure 1.2: Schematic representation of neurulation.....                                                                                                                                   | 18 |
| Figure 1.3: Dorso-ventral patterning in the neural tube.....                                                                                                                               | 19 |
| Figure 1.4: Prosomeric model. ....                                                                                                                                                         | 20 |
| Figure 1.5: Signaling centers of the forebrain and main signaling pathways involved in its<br>patterning.....                                                                              | 23 |
| Figure 1.6: Protein structures of the five ZIC family members. ....                                                                                                                        | 24 |
| Figure 1.7: Expression patterns of the Zic family members during early development.....                                                                                                    | 25 |
| Figure 3.1: Schematic representation of the CRISPR-Cas9 strategy used for deletion and<br>insertion. ....                                                                                  | 38 |
| Figure 3.2: Scheme of pX330A_hCas9_long_chimeric_gRNA_G2P vector.....                                                                                                                      | 39 |
| Figure 3.3: Schematic representation of the generation of the CRISPR-Cas9 expressing vector.....                                                                                           | 40 |
| Figure 4.1: Differentiation of WT hiPSC into AntNPCs.....                                                                                                                                  | 58 |
| Figure 4.2: Differentiation of WT hiPSC into AntNPCs partially recapitulates forebrain patterning.<br>.....                                                                                | 59 |
| Figure 4.3: Generation of ZIC2 <sup>-/-</sup> and ZIC2 <sup>+/-</sup> hiPSC lines with CRISPR-Cas9.....                                                                                    | 61 |
| Figure 4.4: Generation of ZIC2 <sup>FHA/FHA</sup> hiPSC lines as a tool to generate specific ZIC2 binding<br>profiles. ....                                                                | 64 |
| Figure 4.5: Overview of the genomic datasets generated in WT, ZIC2 <sup>FHA/FHA</sup> , ZIC2 <sup>-/-</sup> and ZIC2 <sup>+/-</sup> cell<br>lines. ....                                    | 65 |
| Figure 4.6: The capacity of ZIC2 <sup>-/-</sup> cells to differentiate into neural progenitor cells does not seem<br>to be impaired.....                                                   | 67 |
| Figure 4.7: Heatmaps of the differentially expressed genes. ....                                                                                                                           | 69 |
| Figure 4.8: ZIC2 <sup>-/-</sup> neural progenitors showed increased expression of rostral telencephalon<br>markers and reduced levels of WNT signaling genes. ....                         | 70 |
| Figure 4.9: Expression dynamics of key target genes in WT and ZIC2 <sup>-/-</sup> cells.....                                                                                               | 71 |
| Figure 4.10: ZIC genes are up-regulated in ZIC2 <sup>-/-</sup> neural progenitors. ....                                                                                                    | 73 |
| Figure 4.11: ZIC2 binds both proximal and distal regions of the genome. ....                                                                                                               | 75 |
| Figure 4.12: Differential binding analysis of H3K27ac in ZIC2 <sup>-/-</sup> vs WT cells throughout AntNPC<br>differentiation. ....                                                        | 76 |
| Figure 4.13: Regions gaining H3K27ac in ZIC2 <sup>-/-</sup> cells. ....                                                                                                                    | 78 |
| Figure 4.14: ZIC2-bound regions gaining H3K27ac in ZIC2 <sup>-/-</sup> cells are associated with genes<br>whose silencing upon pluripotency exit is delayed ZIC2 <sup>-/-</sup> cells..... | 80 |
| Figure 4.15: Regions losing H3K27ac mark in ZIC2 <sup>-/-</sup> cells.....                                                                                                                 | 82 |
| Figure 4.16: ZIC2-bound regions losing H3K27ac in ZIC2 <sup>-/-</sup> cells are associated with genes whose<br>expression failed to be induced in ZIC2 <sup>-/-</sup> AntNPC. ....         | 84 |
| Figure 4.17: ZIC2 <sup>-/-</sup> neural progenitor cells show a shift towards the most anterior neural<br>progenitor identities. ....                                                      | 86 |
| Figure 4.18: SAG treatment in WT AntNPC leads to a loss of dorsal identities in favor of more<br>ventral ones. ....                                                                        | 89 |
| Figure 4.19: ZIC2 <sup>-/-</sup> neural progenitors are sensitized to SAG treatment. ....                                                                                                  | 90 |
| Figure 4.20: Dot plot of expression of key neural patterning genes upon SAG treatment.....                                                                                                 | 91 |
| Figure 4.21: ZIC2 <sup>+/-</sup> neural progenitors are sensitive to reduced dosage of ZIC2. ....                                                                                          | 94 |
| Figure 4.22: Future experiments exploring the interplay between ZIC2 and WNT signaling.....                                                                                                | 95 |
| Figure 4.23: Deletion of predicted ZIC2-bound enhancers associated with expression of major<br>anterior neural patterning genes. ....                                                      | 96 |

## List of tables

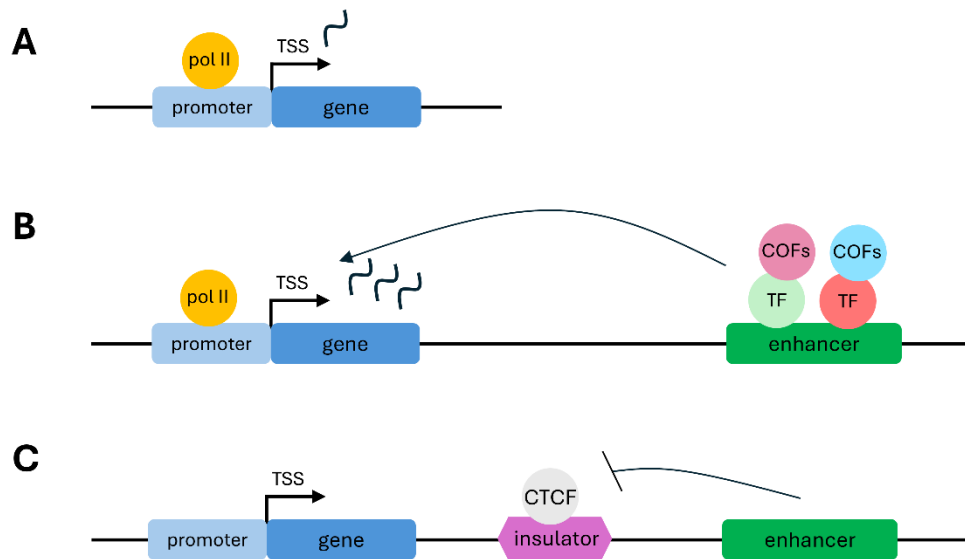
|                                                                                                |    |
|------------------------------------------------------------------------------------------------|----|
| Table 3.1: List of equipment.....                                                              | 31 |
| Table 3.2: Cell culture reagents.....                                                          | 32 |
| Table 3.3: Cell culture media .....                                                            | 33 |
| Table 3.4: Commercial kits .....                                                               | 34 |
| Table 3.5: List of reagents .....                                                              | 34 |
| Table 3.6: List of enzymes.....                                                                | 36 |
| Table 3.7: List of antibodies .....                                                            | 36 |
| Table 3.8: Bacterial reagents.....                                                             | 36 |
| Table 3.9: Composition of the differentiation media of hiPSC into neural progenitor cells..... | 37 |
| Table 3.10: List of gRNAs .....                                                                | 39 |
| Table 3.11: BbsI digestion of the CRISPR-Cas9 vector .....                                     | 41 |
| Table 3.12: List of buffers and composition .....                                              | 42 |
| Table 3.13: cDNA reaction mix .....                                                            | 44 |
| Table 3.14: cDNA synthesis .....                                                               | 45 |
| Table 3.15: Q5 reaction mix .....                                                              | 46 |
| Table 3.16: Q5 Thermocycling program .....                                                     | 47 |
| Table 3.17: OneTaq reaction mix.....                                                           | 47 |
| Table 3.18: OneTaq Thermocycling program.....                                                  | 47 |
| Table 3.19: NZYTaq II reaction mix.....                                                        | 48 |
| Table 3.20: NZYTaq II Thermocycling program.....                                               | 48 |
| Table 3.21: Quantitative PCR reaction .....                                                    | 48 |
| Table 3.22: Quantitative PCR program .....                                                     | 48 |
| Table 3.23: List of primers.....                                                               | 49 |
| Table 3.24: List of software .....                                                             | 51 |

# 1. Introduction

### 1.1. Transcription factors are master regulators of gene expression

From a single cell to a complex organism, the precise spatiotemporal control of gene expression is essential. Thanks to their ability to recognize and bind specific cis-regulatory elements (CREs) of the DNA (i.e., promoters, enhancers, insulators and silencers), transcription factors (TFs) (Lambert et al., 2018; Spitz & Furlong, 2012) play a critical role in the establishment of transcriptional programs that drive the differentiation and function of the multitude of cell types that compose an organism.

Several types of CREs have been identified (Figure 1.1): (i) Promoters are DNA regions localized in proximity to the transcription start sites (TSS) of their genes (Haberle & Stark, 2018). They serve as binding platforms for the recruitment of the basal transcription machinery composed of general transcription factors and RNA polymerase II, allowing basal gene transcription to start. (ii) Enhancers (Long et al., 2016) are defined as distal regulatory elements able to increase the transcription of their target genes via the recruitment of TFs to transcription factor binding sites (TFBS), which in turn interact with co-factors. The three-dimensional conformation of the genome allows enhancers to come into physical proximity with the promoters of their target genes (Bonev et al., 2017; Hsieh et al., 2020; Hua et al., 2021). Epigenetic modifications (histone methylation, acetylation) and binding of certain transcription factors and co-factors defined different types of enhancers: active, primed and poised (Heintzman et al., 2009; Rada-Iglesias et al., 2011; Thurman et al., 2012). Briefly, active enhancers are characterized by an open chromatin state and histone modifications such as H3K27ac and H3K4me1 together with RNA polymerase II and production of eRNA; primed enhancers show H3K4me1 without H3K27ac, indicating an intermediate state ready for activation but not yet active; poised enhancer (or CG-rich enhancers) are bookmarked in pluripotent cells with active features (binding of transcription factors and co-activators like P300, high accessibility and H3K4me1) but lack H3K27ac and they are instead bound by Polycomb proteins that are associated with the deposition of the repressive mark H3K27me3 (Cruz-Molina et al., 2017), reflecting a bivalent state ready to be activated or repressed depending on developmental cues. (iii) Insulators act as boundary elements preventing enhancers from communicating with non-target genes by recruiting architectural proteins (i.e., CTCF) (Özdemir & Gambetta, 2019). (iv) In an analogous manner to enhancers, silencers recruit co-repressors leading to down-regulation of gene expression (Segert et al., 2021).



**Figure 1.1: Schematic representation of the main types of cis-regulatory elements.**

(A) The promoter region located next to the TSS of its gene recruits the general transcription factors together with the RNA polymerase II (pol II), to ensure basal transcription. (B) Enhancers can recruit TFs which in turn interact with co-factors (COFs) to increase the expression of their target genes. (C) Insulators prevent ectopic gene expression by blocking the communication between an enhancer and a gene. They are bound by architectural proteins such as CTCF.

TFs have therefore been defined as master regulators of gene expression, shaping cellular identities, maintaining homeostasis and responding to developmental and environmental cues. The binding motif of a TF is usually a short, between 6 to 12bp, and degenerate DNA sequence, raising questions about its specificity (Spitz & Furlong, 2012). The binding of a TF to DNA is a highly dynamic process involving cooperation between TFs and co-factors with different combinations leading to different transcriptional outputs, answering the issue of specificity. Nevertheless, decoding the cis-regulatory code is not straightforward as several layers of complexity are adding up when interpreting transcriptional changes. Indeed, gene expression is highly context-dependent, influenced by the presence and/or absence of co-factors, by chromatin accessibility and nucleosomes spacing, by post-translational modifications of the TFs themselves, co-factors and chromatin (S. Kim & Wysocka, 2023) and by the rules determining enhancer-promoter communication such as the distance, the biochemical compatibility (Pachano et al., 2021) or the presence of insulators. Overall, gene expression results from the integration of multiple factors, but many of which remain to be elucidated.

To understand how TFs shape cellular identities, we must also consider two inherent features of TFs, redundancy and pleiotropy. Functional redundancy refers to TFs that are able to perform similar functions: a single mutant will not show phenotypic defects while a double mutant will be

affected. Note that redundancy can be complete (a TF compensating fully for the loss of another TF) or partial (a mutation in a single TF can have deleterious defects that are enhanced if a second TF is also mutated). Therefore, TF redundancy plays an important role in ensuring transcriptional robustness (i.e., the ability of the system to remain unchanged under perturbations). On the other hand, pleiotropy is the phenomenon by which a single TF contributes to apparently unrelated phenotypic traits. This occurs because the same TF can be expressed across a broad range of tissues and thus, regulating diverse biological processes.

## **1.2. Transcription factors and disease**

Given their critical role in controlling gene expression and defining cellular identities, it is not surprising that TFs and their regulated regions are frequently linked to human disease (Claringbould & Zaugg, 2021; T. I. Lee & Young, 2013; Maurano et al., 2012; Rada-Iglesias, 2014; R. van der Lee et al., 2020). Despite the precise regulation of transcription during development, TFs are often found to be dosage sensitive (Naqvi et al., 2023; Seidman & Seidman, 2002; R. van der Lee et al., 2020), ranging from phenotypic variations to developmental disorders due to haploinsufficiency or triplosensitivity.

Mouse models have offered significant insights into the study of developmental disorders, providing valuable data on gene regulation. However, several differences between mice and humans can be identified, including morphological differences, variations in developmental timing or in dosage sensitivity. There are numerous examples of developmental genes for which in humans but not mice are haploinsufficient (e.g., *LMX1B*, *TFAP2A*, *SHH*) (Haro et al., 2021; Zhang et al., 1996), emphasizing the importance of also studying developmental diseases within a human-specific context. Gene dosage sensitivity can be exacerbated by gene-environment interactions, where environmental stress increases penetrance and/or expressivity. Indeed, several developmental disorders have been associated with exposure to environmental teratogenic insults (e.g., retinoic acid, alcohol, nicotine, etc.) (Lovely et al., 2017; Nishimura & Kurosawa, 2022).

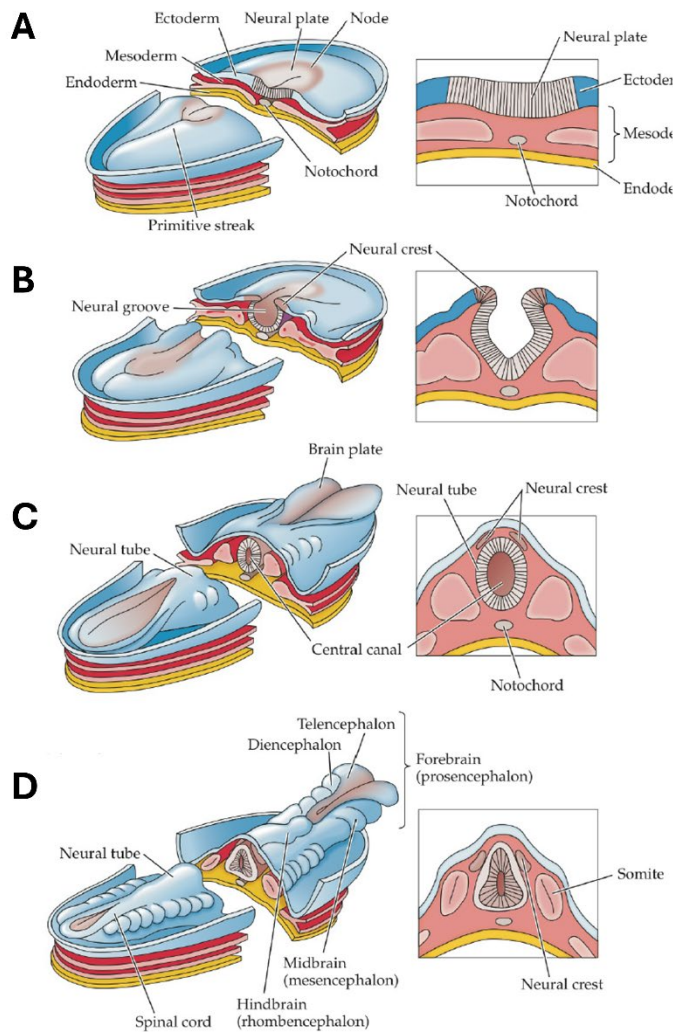
Within this framework, this thesis will focus on *ZIC2* function, a TF from the ZIC family, which has been implicated in holoprosencephaly (HPE), a congenital disorder affecting brain development.

### 1.3. Early brain development

After fertilization, through a series of cleavage divisions, the zygote will form the morula which consists of a compact ball of cells. The morula will reorganize into a blastula, a sphere of cells surrounding an inner fluid-filled cavity: the outer layer of cells of the blastocyst is called the trophoblast and will give rise to extra-embryonic structures, while the inner cell mass will form the embryo proper (Mu et al., 2022). As early embryonic development progresses, the inner cell mass will reorganize into a two-layered disc composed of the hypoblast ventrally and epiblast dorsally, this bilaminar disc sets the dorsal-ventral axis. Gastrulation starts with the appearance of the primitive streak on the surface of the epiblast, setting the antero-posterior axis of the embryo. Cells of the epiblast undergo an epithelial-to-mesenchymal transition and migrate through the primitive streak to form first the innermost germ layer, the endoderm, followed by the mesoderm, whereas, the ectoderm forms the external germ layer, that is responsible for three major roles: specification of the neural plate, the epidermis and the presumptive neural crest (Rossant & Tam, 2022).

The neural plate, formed by the thickening of the ectodermal cells on the dorsal surface of the embryo, is the precursor of the neural tube, which in turn will give rise to the central nervous system. First, the edges of the neural plate will thicken and move upward to form the neural folds, letting apparent the neural groove in a “U-shape”. Progressively, the neural folds will meet at the middle and fuse to form the neural tube. This process during which the neural system is formed is called neurulation (Figure 1.2). On a related note, the neural folds give rise to the neural crest cells, often referred to as the “fourth germ layer”, as they contribute to craniofacial structures, peripheral nerves and various other cell types.

Even before the closure of the posterior part of the neural tube, the anterior part changes drastically forming three primary vesicles: the prosencephalon (forebrain), mesencephalon (midbrain) and the rhombencephalon (hindbrain) along the antero-posterior axis, respectively. The primary vesicles further divide into the secondary vesicles, the forebrain becomes the telencephalon and diencephalon while the rhombencephalon subdivides into the metencephalon and myelencephalon. These vesicles will differentiate into the distinct functional units of the brain. For instance, the telencephalon will give rise to the cerebral hemispheres, olfactory lobes and hippocampus while the diencephalon is responsible for the formation of the optic vesicles, thalamus and hypothalamus. The most posterior part of the neural tube will form the spinal cord.



**Figure 1.2: Schematic representation of neurulation.**

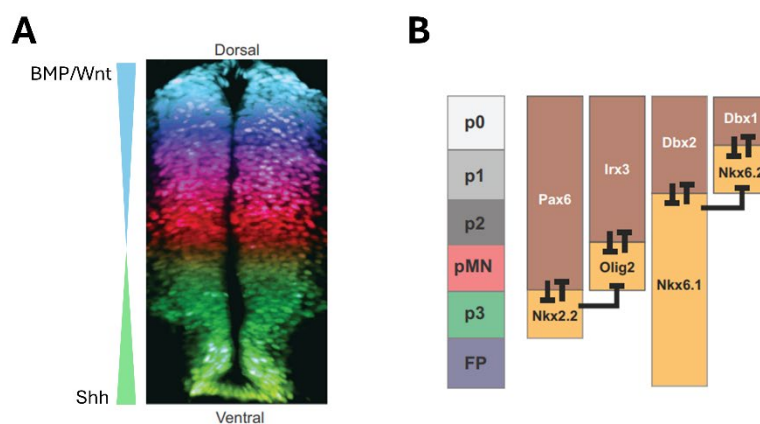
(A) Neural plate (B) Neural groove: the edges of the neural plate thicken and move upwards, forming the neural folds. (C) Closure of the neural tube: the neural folds meet at the middle and fuse, giving rise to the neural tube. (D) Regionalization of the neural tube into the forebrain, midbrain, hindbrain and spinal cord. Figure from Breedlove, 2017.

#### 1.4. Early brain patterning

To understand early brain patterning, it is essential to comprehend how morphogen signaling is transduced and interpreted by the cells to express the correct set of TFs, subsequently defining the developmental fate. The “French-flag” model of Wolpert (Wolpert, 1969) describes how a gradient of morphogens gives information to the cells about their positional identities within a tissue: a morphogen is secreted from a group of cells, usually referred to as organizers or signaling centers and, through diffusion, is distributed within the tissue. Depending on their

distance from the source, the cells will receive different concentrations of morphogens and integrate the signal differentially. Cells that are close to the source will receive a higher concentration of morphogens than the cells located further away. It is important to note that not only the concentration matter but also the time of exposure. In addition, the integration of the signal is also dependent on the TFs already expressed in the cells, meaning that the morphogen by itself does not instruct cellular identities but is rather integrated within a prepatterned tissue (Briscoe & Small, 2015; Sagner & Briscoe, 2017).

Nevertheless, this graded information needs to be converted into sharp boundaries: cross-repressive interactions of TFs (Delás & Briscoe, 2020) ensure the establishment of the proper gene expression program within a tissue by mutually inhibiting the expression of other TFs and thus preventing differentiation into alternative cell fates (Figure 1.3). In other words, very often in development, a signal promoting one cell fate will inhibit the specification into an alternative cell type. This is well illustrated by the example of SHH, a signaling protein emanating ventrally from the notochord and from the floor plate of the neural tube, that determines the ventral progenitor domains within the developing spinal tube, p3, pMN and p2 which are characterized by the expression of *Nkx2.2*, *Olig2* and *Pax6*, respectively. Notably, the loss of *Olig2* and/or *Pax6* leads to the expansion of the *Nkx2.2* domain (Balaskas et al., 2012), reflecting the importance of cross-repression among TFs to determine the proper progenitor domains and subsequently the production of the proper neuron subtypes.

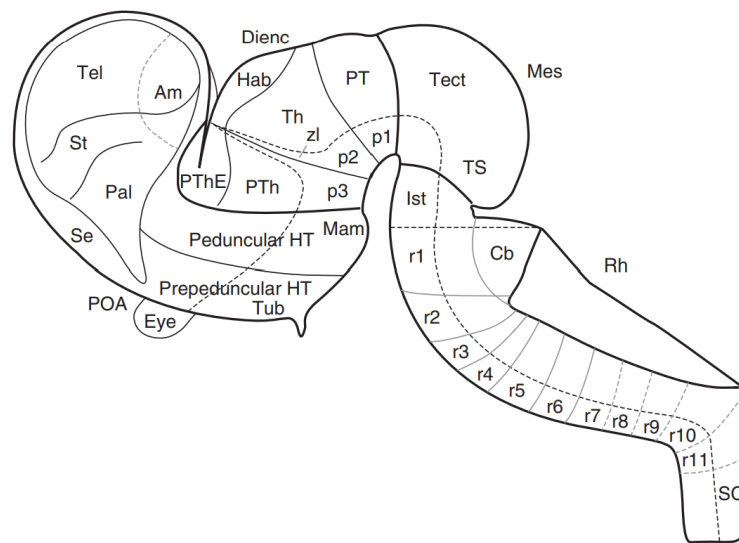


**Figure 1.3: Dorso-ventral patterning in the neural tube.**

(A) Progenitor domain identities along the dorsal-ventral axis are determined by the combinatorial expression of TFs, in response to Shh gradient emanating ventrally and to BMP and Wnt signaling dorsally. (B) TFs expressed in neighboring domains cross-repress each other. Figure from Briscoe & Small, 2015.

The process of brain patterning is highly complex, combining signaling molecules and transcriptional programs to shape the neural tube into the distinct specialized regions that will compose the central nervous system. Here, I will outline some key players of brain patterning, focusing on the forebrain, namely the telencephalon and diencephalon.

Based on gene expression patterns and morphological data, the prosomeric model (Puelles & Rubenstein, 2003) offers a framework for interpreting complex genetic patterns of the developing neural tube. It describes the brain of vertebrates as transversal subunits called neuromeres, each defined by a unique combination of TFs and signaling pathways (Figure 1.4).



**Figure 1.4: Prosomeric model.**

*Illustration of the prosomeric model. It describes the developing brain as transversal subunits, called neuromeres. They are defined based on expression data of TFs and signaling molecules. Am: amygdala; Cb: cerebellum; Dienc: diencephalon; Hab: habenula; HT: hypothalamus; Ist: isthmus; Mam: mammillary; Mes: mesencephalic alar plate; p1-3: prosomeres p1 to p3; Pal: pallidum; POA: pre-optic area; PT: pretectum; PTh: prethalamus; PThE: prethalamic eminence; r1-11: rhombomeres r1 to r11; Rb: rhombencephalon; SC: spinal cord; Se: Septum; St: Striatum; Tect: tectum; Tel: telencephalon; Th: thalamus; TS: torus semicircularis; Tub: tuberal; zl: zona limitans. Figure from Puelles, 2009.*

In development, a recurring set of signaling molecules are involved, namely, WNTs, BMPs, FGFs and SHH (Figure 1.5). The telencephalon located at the anterior end of the neural tube is subjected to *Fgf8* signaling emanating from the anterior neural ridge (ANR) and to *Shh* coming from the ventral midline (floor plate). The interplay between *Fgf* and *Shh* signaling, together with the expression of TFs such as *Gli3*, *Foxg1* or *Pax6* among others, will pattern the telencephalon into distinct proliferative zones. Ventral and dorsal telencephalon is defined by the cross-repression of *Gli3* and *Shh*. *Gli3* is a TF that exists in two forms: the repressor form Gli3R promotes

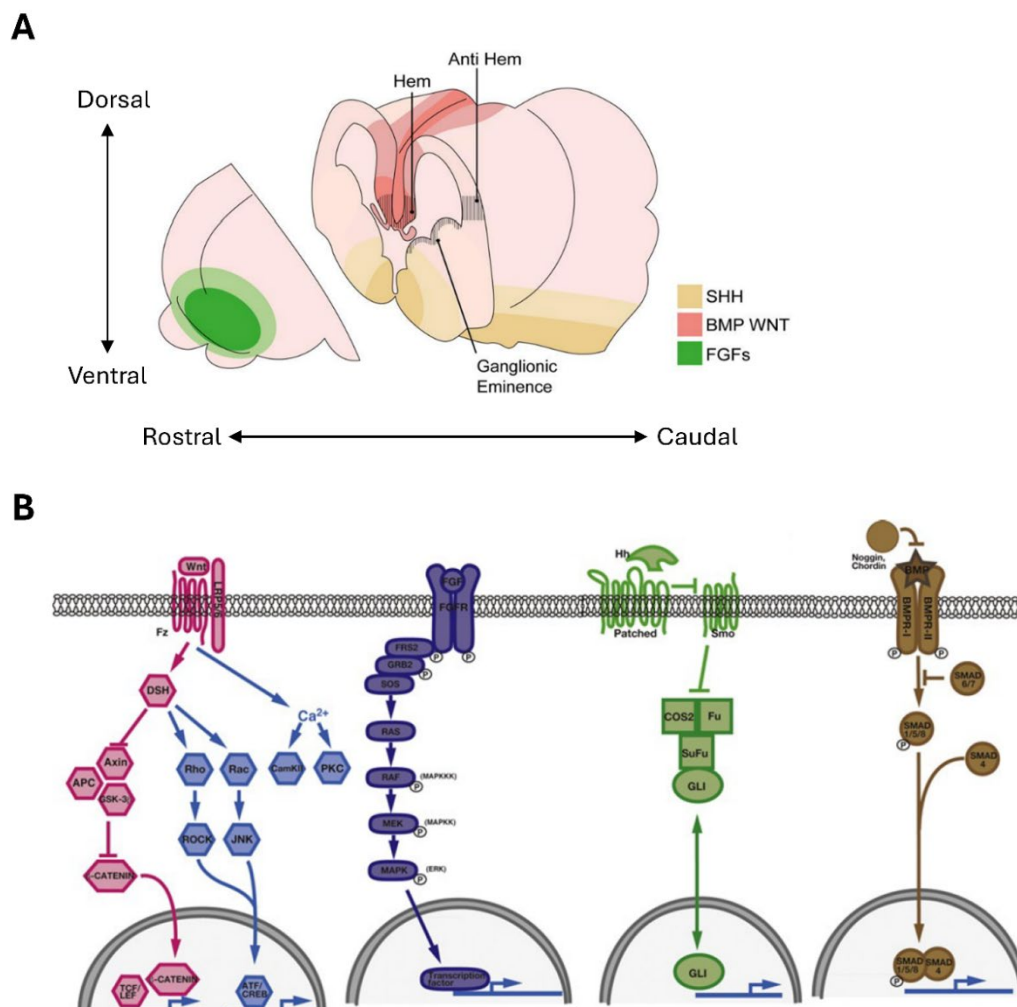
dorsalization of the telencephalon, while the presence of Hedgehog signaling (Hh) ventrally inhibits the post-translational processing of Gli3 into the Gli3R, thus preventing dorsalization (Rallu et al., 2002). Indeed, the loss of *Shh* in mice leads to a loss of ventral identities that is rescued in the double mutant *Shh/Gli3*, indicating that rather than directly promoting ventral identities within the telencephalon, *Shh* inhibits the dorsalization action of Gli3R (Hébert & Fishell, 2008a; Rallu et al., 2002). At most posterior positions of the forebrain (e.g., caudal telencephalon, diencephalon), the ventralization of *Shh* is mediated via the direct activation of the class II genes (e.g., *Nkx2-1*, *Nkx2-2*, *Olig2*) via Gli1 and Gli2 TFs activation (Briscoe et al., 2000; Corbin et al., 2003; Fuccillo et al., 2004). *Pax6*, is another TF involved in dorsal-ventral patterning of the telencephalon. Similarly to Gli3 mutant, the loss of *Pax6* in *Shh*<sup>-/-</sup> mice results in partial rescuing of the ventral identities (Fuccillo et al., 2006).

Other examples of TFs marking the rostral-most identities of the telencephalon are *Six3*, *Foxg1* or *Lhx2* for instance (Suda et al., 2001; Tétreault et al., 2009; Tian et al., 2002). In particular, *Six3*-null embryos lack the telencephalon and the expression domain of *Wnt1* expands in the anterior neuroectoderm (Lagutin et al., 2003; Lavado et al., 2008). Additionally, *Six3* have been shown to promote *Shh* expression in the ventral telencephalon (Geng et al., 2008, 2016; Jeong et al., 2008). Thus, by promoting *Shh* expression while inhibiting WNT expression, *Six3* is essential for the proper patterning of forebrain along both axes, dorsal-ventral and rostro-caudal.

Located at the telencephalic dorsal midline, the cortical hem, is an important signaling center that produces WNT and BMP signaling molecules (Iskusnykh et al., 2023; Shimogori et al., 2004). Wnt ligands are secreted glycoproteins essential for establishing regional identity in the developing forebrain. Indeed, the genetic ablation of the cortical hem in mice (i.e., loss of its major secreted molecule *Wnt3a*), lead to a loss of the hippocampus, region of the brain involved in learning and memory (Caronia-Brown et al., 2014). The loss of the cortical hem affected both dorsal-ventral and rostro-caudal patterning axis (i.e., the dorsomedial neocortex was reduced in size while the ventral cortex was expanded and, the rostral regions of the neocortex were expanded at the expense of the caudal ones). On the other hand, mutations in the BMP pathway in mice (Fernandes et al., 2007) led to a reduced or absent expression of *Wnt2b* and *Wnt3a* throughout the rostro-caudal axis. Moreover, expression of *Msx1*, normally expressed in the cortical hem was lost. BMP signaling is therefore necessary for the proper formation of the cortical hem. Several other TFs have been associated with the cortical hem. For instance, the loss of *Lmx1a* also results in reduced Wnt signaling and cortical hem abnormalities (Iskusnykh et al., 2023), whereas the loss of *Lhx2* or *Foxg1*, originally expressed in the telencephalon result in an expanded cortical hem (Bulchand et al., 2000; Mangale et al., 2008; Muzio & Mallamaci, 2005).

All the previous examples therefore illustrate the crosstalk between dorsal-ventral and rostro-caudal patterning of the developing forebrain.

Within the diencephalon, at the boundary located between the prosomeres p2 and p3, the *zona limitans intrathalamica* (ZLI) is a source of Shh signaling that is expanding dorsally, involved in the patterning of the adjacent regions (Lim & Golden, 2007; Puelles & Martinez, 2013). Examples of TFs marking the diencephalon include *Barhl1* or *Barhl2* (Bulfone et al., 2000; Parish et al., 2016). In figure 1.5 are described the WNT, FGF, SHH and BMP signaling pathways. Interestingly, SHH and WNT/BMP signaling exhibit an antagonistic role regarding forebrain patterning, promoting ventral or dorsal identities, respectively. Effectors of the hedgehog pathway, namely GLI TFs, cooperate with SoxB1 proteins (i.e., Sox1-3 transcriptional activators) to promote ventral identities (Oosterveen et al., 2012). In the case of the canonical WNT signaling pathway, it is unclear whether other TFs cooperate with the  $\beta$ -catenin complex to set up dorsal midline identities.



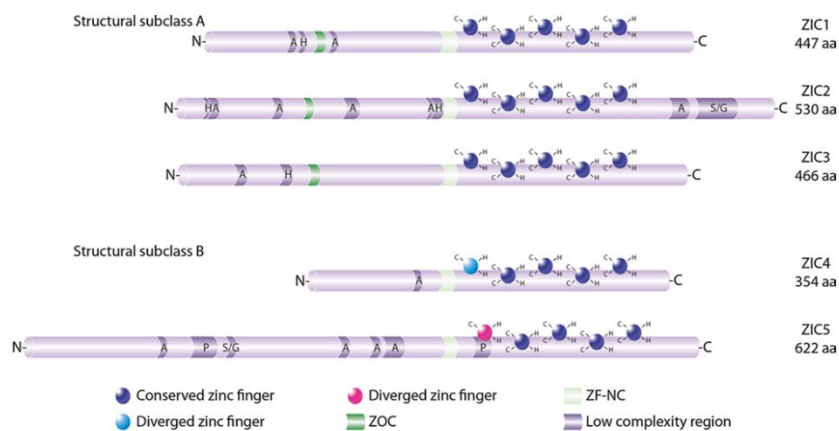
**Figure 1.5: Signaling centers of the forebrain and main signaling pathways involved in its patterning.**

(A) Within the forebrain, formed by the telencephalon and diencephalon, several signaling centers are the source of key morphogens. BMP and WNT (red) are expressed dorsally from the cortical hem and roof plate, SHH emanate ventrally (yellow) from the floor plate and at the most anterior part of the telencephalon, the anterior neural ridge is source of FGF signaling (green). Figure from Sapir et al., 2022. (B) WNT, FGF, SHH and BMP signaling pathways. WNT ligands bind to the frizzled receptor (Fz) and co-receptor LRP5/6 on the cell membrane which in turn recruit dishevelled (DSH) that prevents the destruction complex (Axin, APC, GSK-3) from degrading  $\beta$ -catenin, that can therefore accumulate in the cytoplasm and translocate to the nucleus to activate its target genes in association with TCF/LEF TFs. FGF ligands bind to their receptors FGFR (tyrosine kinase receptors), that results in the dimerization and activation of the receptors and the subsequent activation of the RAS/MAPK pathway. SHH ligands bind to PTCH1 (patched) receptor which remove its repression on SMO, therefore preventing the phosphorylation of GLI TFs that can in turn activate their target genes. Receptors of BMP ligands consist of a complex composed of BMPR type I and type II receptors. The presence of BMP ligands leads to the phosphorylation of the receptors and to the activation of SMAD1, 5 or 8 that can form a complex with SMAD4 to activate the expression of its target genes. Figure from Takagaki et al., 2012.

### 1.5. ZIC family of transcription factors

The ZIC family of TFs is composed of five members, ZIC1 to ZIC5. Several functions have been associated with ZIC proteins, ranging from maintenance of pluripotency, patterning of the neural tube to skeletal and eye development (Houtmeyers et al., 2013; Solomon et al., 2010). ZIC genes are organized in tandem pairs with ZIC2 and ZIC5 being located on chromosome 13 and ZIC1 and ZIC4 on chromosome 3, whereas ZIC3 is located on chromosome X as a singleton. Similar organization in tandem is also found in mice, with *Zic2/Zic5* located on chromosome 14, *Zic1/Zic4* on chromosome 9 and *Zic3* on chromosome X (Houtmeyers et al., 2013). At both DNA and protein levels, ZICs are highly conserved between mouse and human. At the functional level, there are also conserved as mutations in *Zic* genes in mouse recapitulate phenotypes observed in human. All five ZIC proteins contain a highly conserved zinc finger domain that consists of five tandem Cys2His2-type zinc fingers (Figure 1.6), with the exception of the first zinc fingers of ZIC4 and ZIC5 that are divergent. This domain has been linked with protein-protein interactions as well as DNA binding (Pourebrahim et al., 2011). They also share a ZF-NC domain (zinc finger N-flanking conserved), located immediately upstream of the zinc finger domain. ZIC1, ZIC2 and ZIC3 possess a ZOC motif (ZIC opa conserved) that have been shown to be involved in protein-protein interactions (Mizugishi et al., 2004). Depending on the presence of the ZOC motif and the degree of conservation, ZIC proteins have been divided into two structural subclasses, with ZIC1, ZIC2

and ZIC3 being part of the structural subclass A and ZIC4 and ZIC5 part of subclass B (Houtmeyers et al., 2013).



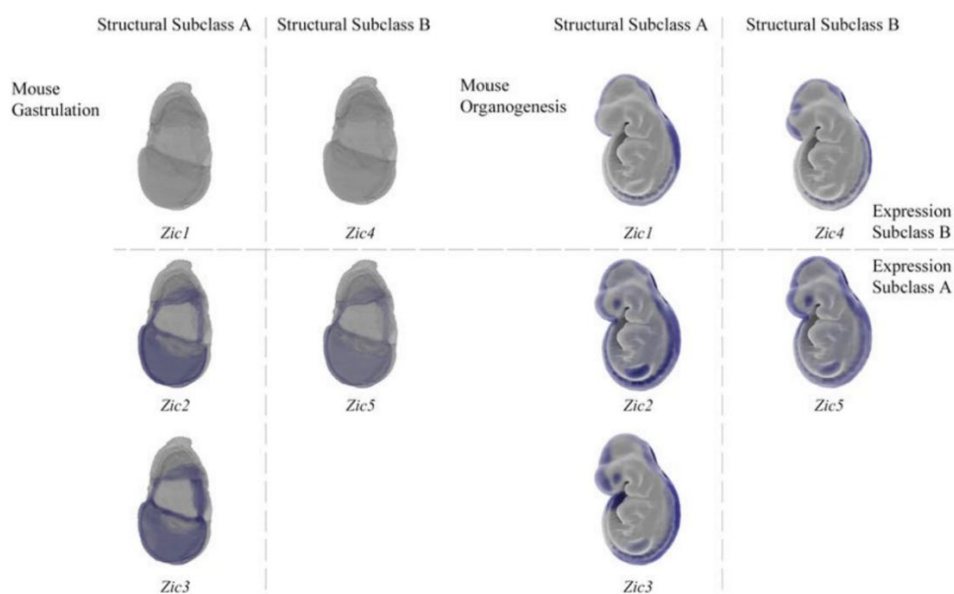
**Figure 1.6: Protein structures of the five ZIC family members.**

All five ZIC proteins share a conserved zinc finger domain that consists of five tandem Cys2His2-type zinc fingers. The first zinc finger of ZIC4 and ZIC5 is divergent. They all share a conserved ZF-NC domain located upstream of the zinc finger domain. Different low complexity regions, enriched in specific amino acids, are also present within the ZIC proteins. Additionally, ZIC1, ZIC2 and ZIC3 possess a ZOC motif. Depending on the degree of conservation of the first zinc finger and the presence of the ZOC motif, ZIC proteins have been classified into structural subclass A (ZIC1, ZIC2 and ZIC3) and structural subclass B (ZIC4 and ZIC5). Figure from Houtmeyers et al., 2013.

Based on their expression patterns during mouse embryogenesis (Figure 1.7), we can separate the *Zic* genes into two groups. *Zic2*, *Zic3* and *Zic5* are the first *Zic* members that are being expressed, starting already before implantation (Brown & Brown, 2009). During gastrulation, there are expressed in the ectoderm and in part of the newly formed mesoderm (Elms et al., 2004; Furushima et al., 2000; Houtmeyers et al., 2013). By the early head-fold stage, their expression becomes restricted to the neuroectoderm and later to the dorsal neuroectoderm (that corresponds to the region that will form the neural crest cells and dorsal neurons) and to the pre-somitic and somatic regions of the lateral mesoderm (Diamand et al., 2018). Expression of *Zic1* and *Zic4* is first detected during early organogenesis in the neuroectoderm and somitic mesoderm. As development progresses (E9.5), all *Zic* genes are co-expressed in the dorsal spinal cord, neural tube and in the somites except for *Zic4* not being expressed in the roof plate of the neural tube (Elms et al., 2004; Furushima et al., 2000; Nagai et al., 1997). At E10.5, the *Zic* genes of the first group (*Zic2*, *Zic3* and *Zic5*) are detected in the developing eye and limb buds (Diamand et al., 2018; Nagai et al., 1997). At later stages (E12.5), *Zic* expression in the brain is restricted to the dorsal midline structures (cortical hem, septum and ventricular and subventricular zone) and to

the ventral part of the neural tube (Diamand et al., 2018). Interestingly, the expression pattern of each group mirrors their genomic arrangement, possibly due to shared regulatory regions (Houtmeyers et al., 2013).

The *ZIC* family of TFs illustrates well the two concepts of pleiotropy and redundancy mentioned in previous sections. Indeed, the phenotypes related to *Zic2* loss of function are not limited to HPE, as it has also been associated with neural tube closure defects (spina bifida), heart developmental defects and skeletal abnormalities (Nagai et al., 2000). On the other hand, partial redundancy is illustrated by the compound mutant of *Zic2* and *Zic3* (Inoue et al., 2007). Both single mutants show neural tube defects and skeletal defects that are exacerbated in the double mutant.



**Figure 1.7: Expression patterns of the *Zic* family members during early development.**

*During gastrulation, only Zic2, Zic3 and Zic5 are expressed in the ectoderm and mesoderm, whereas expression of Zic1 and Zic4 is initiated during organogenesis. At this later stage, Zic expression is limited to the central nervous system. Figure from Houtmeyers et al., 2013.*

## 1.6. ZIC2-related HPE

Insight into ZIC2 function during brain development is given by the different mouse models of *Zic2*-related HPE. HPE is a congenital disorder affecting the dorsal-ventral forebrain axis, characterized by a lack of separation of the two brain hemispheres. It is the most common structural defect of the human forebrain, affecting 1 in 250 conceptuses and 1 in 16000 live births (Dubourg et al., 2007; Tekendo-Ngongang et al., 1993). Phenotypic expression of this disorder is highly variable and defined by incomplete penetrance. Classical HPE can be classified into three categories depending on the level of severity (alobar, semilobar and lobar, from more to less

severe, respectively). In classical HPE, defects are mainly localized ventrally. The middle interhemispheric variant (MIHV), a less severe form of HPE, is characterized by defects localized dorsally. Several genes have been associated with this disorder, among them *SHH*, *ZIC2*, *SIX3*, *TGIF1* or *GLI2* for instance, *ZIC2* being the second most mutated gene in HPE. Interestingly, while all these genes have been associated with classical HPE, only *ZIC2* have also been associated with the MIVH form. Nevertheless, the majority of the patients (> 95%) with *ZIC2*-related HPE have the classical form (Barratt & Arkell, 2018). These genes are haploinsufficient, meaning that the loss of function in one allele is enough to lead to the development of HPE (Dubourg et al., 2007; Tekendo-Ngongang et al., 1993). In addition to genetic factors, exposure to teratogens, such as maternal alcohol consumption or to retinoic acid, have been associated with HPE etiology (Hong & Krauss, 2017; Michael Cohen & Shiota, 2002), highlighting the complexity of HPE and the multiple processes affected.

In mouse, the severe loss of function of *ZIC2* (Kumba allele) results in classical HPE (Warr et al., 2008), indicating that *ZIC2* plays a role in ventral patterning. It has been shown that during gastrulation, *ZIC2* interacts with the NODAL pathway during the establishment of the anterior notochord. Loss of *ZIC2* at this stage disrupts the proper establishment of the prechordal plate which in turn prevents initiation of *Shh* signaling, leading to ventral patterning defects resembling classical HPE (Warr et al., 2008). Conversely, mouse models with a partial reduction (approximately 80%) in *ZIC2* activity result in MIHV (Nagai et al., 2000), where the forebrain defects are localized dorsally. In these mice, roof plate formation is impaired, leading to absent or defective dorsal structures. Yet, the exact molecular mechanism by which *ZIC2* might control the development of dorsal structures remains unknown. Interestingly, dorsal patterning seems to be more sensitive to *ZIC2* dosage compared to ventral patterning, as ventral forebrain signaling does not seem to be impaired in *Zic2* hypomorphic mutants where *ZIC2* expression/activity is reduced to ~20% (Nagai et al., 2000). In summary, *Zic2* plays a dual role in forebrain patterning, being necessary for the proper establishment of ventral identities through the interaction with the NODAL pathway and in the establishment of dorsal identities during neurulation, where it is expressed in the dorsal midline of the neural tube.

### **1.7. *ZIC2* as an enhancer binding factor**

At the mechanistic level, little is known about *ZIC2* target genes during neural induction. Nevertheless, different studies depict *ZIC2* as an essential enhancer binding factor in pluripotency. *ZIC2* ChIP-seq in mESC (Luo et al., 2015), revealed that *ZIC2* binds to enhancers,

including active and poised ones. They also show that ZIC2 interacts with the chromatin remodeler complex NuRD. Studies in drosophila (Soluri et al., 2020) have shown that *Opa* gene (homolog of the *Zic* genes) facilitates chromatin opening. ChIP-seq analysis of pluripotency network in mESC and EpiSC, revealed that the dominance of SOX2 and OCT4 in mESC in terms of the number of binding sites and bound CREs is shifted toward a dominance of ZIC2 and OTX2 in EpiSC (Matsuda et al., 2017). Consistently with these observations, a recent study in human pluripotent stem cells (hESC) showed that the loss of both ZIC2/ZIC3, but not of ZIC2 alone, in primed hESC is correlated with a loss of chromatin accessibility at CREs and down-regulation of nearby genes (Hossain et al., 2024). In the same study, they demonstrated that the chromatin opening capacity of ZIC2/ZIC3 is mediated via the recruitment of the SWI/SNF complex. Therefore, these results argue in favor of a pioneering role for ZIC2 and ZIC3 in pluripotent cells. However, these two ZIC factors seem to be redundant in human pluripotent cells and it is currently unclear whether ZIC2 could play a similar and perhaps redundant pioneering role function during neural induction.

### **1.8. Modeling early brain development *in vitro***

Although animal models have provided valuable data, understanding specific aspects of human brain development remains challenging. In this context, human pluripotent stem cells and their differentiation towards a neural fate provide new opportunities to study the underlying molecular mechanisms (Suzuki & Vanderhaeghen, 2015). *In vitro* differentiation offers several advantages as they represent highly tractable systems that allow for easy genetic modifications of hiPSC before differentiation while yielding a large amount of biological material for subsequent molecular applications, such as RNA-seq or ChIP-seq for instance.

Neural induction of pluripotent stem cells is considered as the default differentiation pathway in the absence of signal for self-renewal. This default pathway is triggered by the inhibition of TGF $\beta$  and BMP signaling. One other main advantage of *in vitro* differentiation system is the possibility to modulate the cellular identities obtained by the addition of different drugs that either inhibit or activate the chosen pathway. For instance, retinoic acid is known to promote posterior identities of the central nervous system while the addition of a WNT inhibitor will favor the most anterior identities, with the expression of the TF *FOXP1* (Chambers et al., 2009). Conversely, the addition of a WNT agonist will promote midbrain identities. In this work, we took advantage of human induced pluripotent stem cells (hiPSC) and their differentiation into anterior neural progenitor cells (AntNPC), following the protocol from (Tchieu et al., 2017). This differentiation protocol is

based on dual SMAD inhibition, using SB and LDN molecules that inhibits TGF $\beta$  and BMP signaling pathways, respectively.

### **1.9. Genomics to study ZIC2 function**

In this thesis, we employed RNA-seq and ChIP-seq techniques to study the role of *ZIC2* during the induction of anterior neural identities. They provided valuable results in the understanding of *ZIC2* function, interrogating the transcriptional changes upon loss of *ZIC2*, mapping *ZIC2* binding sites across differentiation and investigating key changes in chromatin activity. Nevertheless, interpretation of bulk measurements can be challenging as they produce an average readout, failing to capture cell population heterogeneity. For this reason, we also performed single-cell RNA-seq experiments, to uncover the heterogeneity of the cellular identities obtained during the differentiation of hiPSC into anterior neural progenitor cells and better interpret the transcriptional changes observed. Indeed, recent advances in single-cell measurement techniques have provided key advantages in the understanding of biological processes, with the possibility to assess transcriptional states at the single-cell resolution. Despite the new advances, limitations of single cell techniques persist, such as the sparsity of the data or the requirement of advanced computational tools (Lähnemann et al., 2020; Trapnell, 2015). Nevertheless, single cell omics represent a powerful tool for the understanding of cellular heterogeneity and dynamics of gene expression changes. Several brain cell atlases using single-cell transcriptomics from human embryo have/will help identifying cell types and cellular states of the developing human brain (Braun et al., 2023; X. Chen et al., 2024; Y. Li et al., 2023; Xu et al., 2023; Zeng et al., 2023).

## 2. Objectives

The main objectives of this thesis are:

1. To understand the role of *ZIC2* during neural induction and brain patterning in a human context.
2. To elucidate the regulatory network controlled by *ZIC2* during the establishment of anterior neural identities.

To do so, we took advantage of hiPSC and differentiated them into AntNPC. Prior to differentiation, hiPSC have been genetically engineered, using CRISPR-Cas9 technology. Throughout the differentiation, RNA-seq, ChIP-seq and scRNA-seq data have been generated.

### **3. Materials and Methods**

### 3.1. Material

#### 3.1.1. Equipment

Table 3.1: List of equipment

| Equipment                                                            | Company            |
|----------------------------------------------------------------------|--------------------|
| <b>Bacteria incubators</b>                                           |                    |
| Infors HT Multitron Ampere Chart Multitron II                        | Infors HT          |
| HeraTherm Incubator                                                  | Thermo Scientific  |
| <b>Cell culture</b>                                                  |                    |
| Hera Cell 240i CO2 Incubator                                         | Thermo Scientific  |
| Panasonic KM-CC17RH2E CO2 incubator                                  | Panasonic          |
| TC20™ Automated Cell Counter                                         | BioRad             |
| Telstar Bio II Advance Class II Cabinet                              | Telstar            |
| Nikon Eclipse TS100                                                  | Nikon              |
| ZOE Fluorescent Cell Imager                                          | BioRad             |
| <b>Centrifuges</b>                                                   |                    |
| Eppendorf Centrifuge 5810 R                                          | Eppendorf          |
| Eppendorf Centrifuge 5425 R                                          | Eppendorf          |
| <b>Sonicator</b>                                                     |                    |
| EpiShear™ Probe Sonicator                                            | Active motif       |
| <b>Thermal cyclers</b>                                               |                    |
| T100 Thermal Cycler                                                  | BioRad             |
| C1000 Touch Thermal Cycler                                           | BioRad             |
| SimpliAmp Thermal Cycler                                             | Applied Biosystems |
| Veriti 96 Well Thermal Cycler                                        | Applied Biosystems |
| qPCR Opus CFX                                                        | BioRad             |
| <b>DNA electrophoresis</b>                                           |                    |
| Mini-Sub Cell GT Horizontal Electrophoresis System (7x7 cm)          | BioRad             |
| Wide Mini-Sub Cell GT Horizontal Electrophoresis System (15 x 10 cm) | BioRad             |
| PowerPac™ Basic Power Supply                                         | BioRad             |
| Agilent 4200 TapeStation System                                      | Agilent            |

| Equipment                                        | Company           |
|--------------------------------------------------|-------------------|
| <b>Western Blot</b>                              |                   |
| Mini-PROTEAN Tetra Vertical Electrophoresis Cell | BioRad            |
| Mini-PROTEAN Tetra Companion Running Module      | BioRad            |
| PowerPac Universal Power Supply                  | BioRad            |
| Amersham ImageQuant 800                          | Cytiva            |
| <b>Spectrophotometers/Fluorimeters</b>           |                   |
| Nanodrop 2000 Spectrophotometer                  | Thermo Scientific |
| Qubit 4 fluorometer                              | Invitrogen        |
| Multiskan FC                                     | Thermo Scientific |
| OD600 DiluPhotometer                             | IMPLEN            |
| <b>Thermoblocks</b>                              |                   |
| Thermo Block Eppendorf                           | Eppendorf         |
| Thermo Block Dtabis                              | Ditabis           |

### 3.1.2. Cell lines

Wild-type human induced pluripotent stem cells (hiPSC) were provided by Tomo Saric lab (University of Cologne, Germany) and are available through the European Bank for Induced Pluripotent Stem Cells (EBiSC) (ID: UKKi011-A; <https://ebisc.org/UKKi011-A>). Using CRISPR-Cas9 technology, several cell lines with heterozygous ( $ZIC2^{+/-}$  hiPSC; clones #10, #11 and #28) or homozygous ( $ZIC2^{-/-}$  hiPSC; clones #14, #25 and #34) deletions of the *ZIC2* gene were generated. In addition, deletion of the *ZIC1* gene within  $ZIC2^{-/-}$  background has been generated ( $ZIC1^{-/-}; ZIC2^{-/-}$  hiPSC; clone #13). Moreover, the endogenous *ZIC2* gene was tagged with a Flag-HA sequence in C-terminal ( $ZIC2^{FHA/FHA}$  hiPSC; clones #17 and #5).

### 3.1.3. Cell culture reagents and media

Table 3.2: Cell culture reagents

| Reagent                        | Source                  | Catalog number |
|--------------------------------|-------------------------|----------------|
| Accutase                       | ThermoFisher Scientific | A1110501       |
| Dimetilsulfóxido (DMSO)        | Sigma-Aldrich           | D2650          |
| Fetal Bovine Serum (FBS)       | ThermoFisher Scientific | 10500064       |
| FuGENE HD Transfection Reagent | Promega                 | E2311          |
| Geltrex                        | ThermoFisher Scientific | A1413302       |

| Reagent                  | Source                  | Catalog number |
|--------------------------|-------------------------|----------------|
| LDN193189 (LDN)          | Sigma-Aldrich           | SML0559        |
| Opti-MEM                 | ThermoFisher Scientific | 31985070       |
| Puromycin                | Sigma-Aldrich           | P8833          |
| SB431542 (SB)            | R&D Systems             | 1614           |
| Smoothened Agonist (SAG) | Sigma-Aldrich           | 566660         |
| TrypLE                   | ThermoFisher Scientific | 12-604-021     |
| Y-27632 (ROCKi)          | R&D Systems             | 1254           |

Table 3.3: Cell culture media

| Media                   | Components                       | Volume | Source                  | Catalog number |
|-------------------------|----------------------------------|--------|-------------------------|----------------|
| mTeSR media             | mTeSR1                           | 400mL  | STEMCELL technologies   | 85850          |
|                         | mTeSR supplement 100X            | 100mL  | STEMCELL technologies   | 85850          |
|                         | Antimycotic/antibiotic           | 5mL    | Sigma-Aldrich           | A5955          |
| KSR media               | DMEM KnockOut                    | 410mL  | ThermoFisher Scientific | 10829018       |
|                         | KnockOut Serum Replacement (KSR) | 75mL   | ThermoFisher Scientific | 10828028       |
|                         | Antimycotic/antibiotic           | 5mL    | Sigma-Aldrich           | A5955          |
|                         | L-Glutamine                      | 5mL    | ThermoFisher Scientific | 25030024       |
|                         | MEM NEAA                         | 5mL    | ThermoFisher Scientific | 11140035       |
|                         | 2-mercaptoethanol                | 0.5mL  | ThermoFisher Scientific | 21985023       |
| N2 media                | DMEM/F-12                        | 500mL  | ThermoFisher Scientific | 11320033       |
|                         | 2-mercaptoethanol                | 0.5mL  | ThermoFisher Scientific | 21985023       |
|                         | Sodium Bicarbonate               | 1g     | Sigma-Aldrich           | S5761          |
|                         | D-(+)- Glucose                   | 0.78g  | Roth                    | HN06           |
|                         | Progesterone                     | 10µL   | Sigma-Aldrich           | P8783          |
|                         | N2 supplement                    | 5mL    | R&D Systems             | AR009          |
| Freezing media<br>hiPSC | mTeSR media                      | 40%    | STEMCELL technologies   | 85850          |
|                         | DMSO                             | 20%    | Sigma-Aldrich           | D2650          |
|                         | KSR                              | 40%    | ThermoFisher Scientific | 10828028       |

## 3.1.4. Molecular biology kits, reagents, enzymes and antibodies

Table 3.4: Commercial kits

| <b>Kits</b>                                    | <b>Source</b>           | <b>Catalog number</b> |
|------------------------------------------------|-------------------------|-----------------------|
| BD Rhapsody™ Cartridge Kit                     | BD Biosciences          | 633733                |
| BD Rhapsody™ Cartridge Reagent Kit             | BD Biosciences          | 633731                |
| BD Rhapsody™ cDNA Kit                          | BD Biosciences          | 633773                |
| BD™ Hu Single Cell Sample Multiplexing Kit     | BD Biosciences          | 633781                |
| innuPREP RNA Mini Kit                          | Analytik Jena           | 845-KS-2040010        |
| NZY Tissue gDNA Isolation kit                  | NZYtech                 | MB13503               |
| NZYMiniprep                                    | NZYtech                 | MB01001               |
| Pierce BCA Protein Assay kit                   | ThermoFisher Scientific | 23225                 |
| Protease Inhibitor Cocktail, Complete Ultra    | Sigma-Aldrich           | 5892791001            |
| ProtoScript II First Strand cDNA Synthesis Kit | NEB                     | E6560L                |
| QIAquick PCR & Gel Cleanup Kit                 | Qiagen                  | 28506                 |
| SPEEDTOOLS PCR Clean-Up kit                    | Bitools                 | 21201-4205            |

Table 3.5: List of reagents

| <b>Reagents</b>                              | <b>Source</b> | <b>Catalog number</b> |
|----------------------------------------------|---------------|-----------------------|
| 2-mercaptoethanol                            | ThermoFisher  | 21985023              |
| 30% Acrylamide-Bisacrylamide solution        | Sigma-Aldrich | A3574                 |
| Agarose                                      | NZYTech       | MB02703               |
| Ammonium persulfate (APS)                    | Merck         | A3678                 |
| Bovine Serum Albumin Powder (BSA)            | Sigma-Aldrich | A9085                 |
| Bromophenol Blue                             | Sigma-Aldrich | 114405                |
| Dithiothreitol (DTT)                         | VWR           | 0281                  |
| Dynabeads™ Protein A for Immunoprecipitation | Invitrogen    | 10002D                |
| EDTA                                         | Sigma-Aldrich | E5134                 |
| EGTA                                         | VWR           | 0732                  |
| Ethanol 96% Molecular Biology Grade          | VWR           | C20824.2              |

| Reagents                                     | Source                   | Catalog number |
|----------------------------------------------|--------------------------|----------------|
| Formaldehyde solution 37%                    | Sigma-Aldrich            | 252549         |
| Glycerol bidistilled 99.5%                   | VWR                      | 24.388.295     |
| Glycine powder 99%                           | Sigma-Aldrich            | G8898          |
| HEPES                                        | VWR                      | 30.487.297     |
| Hydrochloric acid (HCl)                      | Sigma-Aldrich            | 320331         |
| Hydrogen Peroxide                            | Sigma-Aldrich            | 216763         |
| Invitrogen Ambion Sodium Acetate (3M), pH5.5 | Thermo Fisher Scientific | AM9740         |
| LB agar                                      | Sigma-Aldrich            | L2897          |
| Lithium Chloride (LiCl)                      | Sigma-Aldrich            | 310468         |
| Lithium Chloride (LiCl) ACS reagent, 99%     | Sigma-Aldrich            | 310468         |
| Luminol 97%                                  | Sigma-Aldrich            | A8511          |
| Magnesium Chloride (MgCl <sub>2</sub> )      | VWR                      | 25.108.295     |
| Methanol                                     | VWR                      | 20.847.360     |
| Na-Deoxycholate                              | Merck                    | D6750          |
| Na-Lauroylsarcosine sodium salt              | Sigma-Aldrich            | L9150          |
| NP-40 Surfact-Amps™                          | Thermo Fisher Scientific | 85124          |
| p-Coumaric acid                              | Sigma-Aldrich            | C9008          |
| Phosphate Buffered Saline (PBS)              | Sigma-Aldrich            | D8662          |
| Potassium Chloride (KCl)                     | Sigma-Aldrich            | P9333          |
| Sodium Chloride (NaCl) ACS reagent, 99.0%    | Sigma-Aldrich            | S9888          |
| Sodium dodecyl sulfate (SDS)                 | Sigma-Aldrich            | L3771          |
| TEMED                                        | VWR                      | 0761           |
| Triton X-100 Molecular Biology Grade         | VWR                      | 437002A        |
| Trizma base                                  | Sigma-Aldrich            | T1503          |
| Tween20                                      | VWR                      | 437082Q        |

Table 3.6: List of enzymes

| Enzymes                                                                       | Source                  | Catalog number |
|-------------------------------------------------------------------------------|-------------------------|----------------|
| BbsI                                                                          | NEB                     | R0539L         |
| cOmplete ULTRA Tablets, Mini, EDTA-free, EASYpack Protease Inhibitor Cocktail | Sigma-Aldrich           | 5892791001     |
| NZYSpeedy qPCR Green Master Mix                                               | NZYtech                 | MB224          |
| NZYTaq II 2x Green Master Mix                                                 | NZYtech                 | MB35803        |
| OneTaq                                                                        | NEB                     | M0480L         |
| Proteinase K                                                                  | ThermoFisher Scientific | EO0492         |
| Q5                                                                            | NEB                     | M0491S         |
| RNAse A                                                                       | ThermoFisher Scientific | EN0531         |
| T4 ligase                                                                     | ThermoFisher Scientific | EL0013         |

Table 3.7: List of antibodies

| Antibody name                                         | Source                  | Catalog number |
|-------------------------------------------------------|-------------------------|----------------|
| Anti-HA tag                                           | Abcam                   | ab9110         |
| Anti-ZIC2                                             | Abcam                   | ab150404       |
| Goat anti-Rabbit IgG (H+L)<br>Secondary Antibody, HRP | ThermoFisher Scientific | 65-6120        |
| Histone H3K27ac antibody                              | Active motif            | 39133          |
| Histone H3K27me3 antibody                             | Active motif            | 36155          |

### 3.1.5. Molecular cloning

Table 3.8: Bacterial reagents

| Name                                 | Description             | Source                      |
|--------------------------------------|-------------------------|-----------------------------|
| pX330-hCas9-long-chimeric-gRNA-GFP   | CRISPR/Cas9 vector      | Dr. Leo Kurian's laboratory |
| TOP10 competent <i>E. coli</i> cells | Thermal competent cells | Dr. Leo Kurian's laboratory |

### 3.2. Methods

#### 3.2.1. Culture of hiPSC

hiPSC were grown on Geltrex-coated plates with mTeSR media supplemented with antibiotic/antimycotic solution. Confluent cells were passaged with StemPro Accutase. After thawing or splitting, the media was supplemented with ROCK inhibitor (Y-27632) for one day. hiPSC were frozen in freezing medium and stored at -140°C. The cells were incubated at 37°C with 5% CO<sub>2</sub>.

#### 3.2.2. Differentiation of hiPSC into anterior neural progenitor cells

hiPSC were differentiated into anterior neural progenitor cells following the protocol from Tchieu et al., 2017, with few changes. In summary, the day prior the differentiation, the cells were plated at a density of 200-300,000 cells/cm<sup>2</sup> on Geltrex-coated plates in mTeSR supplemented with ROCKi. The following days, the differentiation media (Table 3.3) was changed as described in Table 3.9. SB and LDN act as inhibitors of the TGF- $\beta$  and BMP signaling pathways, respectively.

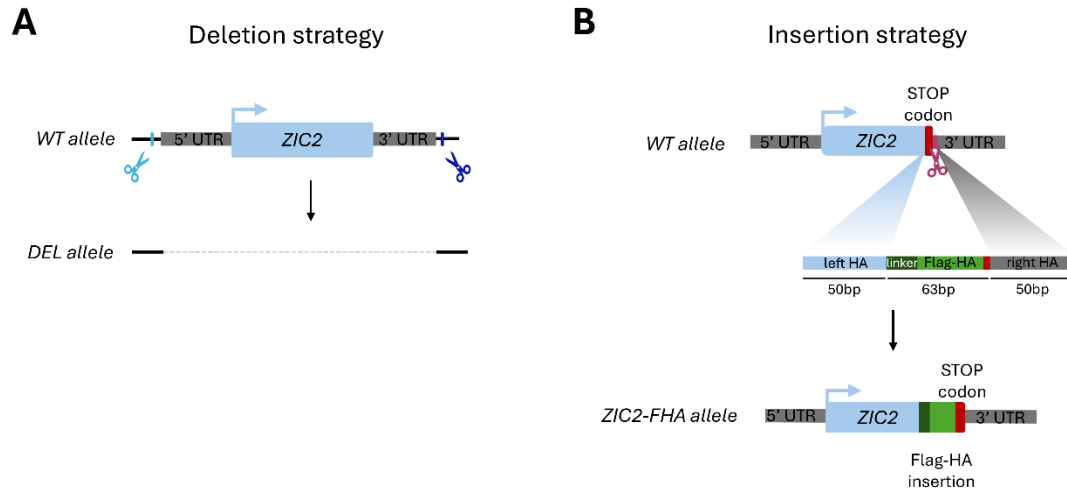
In the case of the SAG-treated differentiation, 0.5 $\mu$ M of SAG were added from day 4 until the end, while control samples were treated with an equivalent volume of DMSO. SAG is an agonist of the hedgehog pathway, via activation of the Smoothened receptor. Based on concentrations commonly used in the literature, two doses of SAG (0.1 $\mu$ M and 0.5 $\mu$ M) were tested. The final dose of 0.5 $\mu$ M was selected because it resulted in greater transcriptional changes, easier to quantify by RT-qPCR.

Table 3.9: Composition of the differentiation media of hiPSC into neural progenitor cells

| Differentiation day | % KSR media | % N2 media | 10 $\mu$ M SB | 500nM LDN |
|---------------------|-------------|------------|---------------|-----------|
| Day 0 – 4           | 100         | 0          | +             | +         |
| Day 4 – 6           | 75          | 25         | +             | -         |
| Day 6 – 8           | 50          | 50         | +             | -         |
| Day 8 – 10          | 25          | 75         | +             | -         |
| Day 10 – 12         | 0           | 100        | +             | -         |

### 3.2.3. Generation of transgenic hiPSC lines with CRISPR-Cas9

Using CRISPR-Cas9 technology, hiPSC were genetically engineered to either delete a locus of interest or to integrate a DNA sequence at a specific location (Figure 3.1).

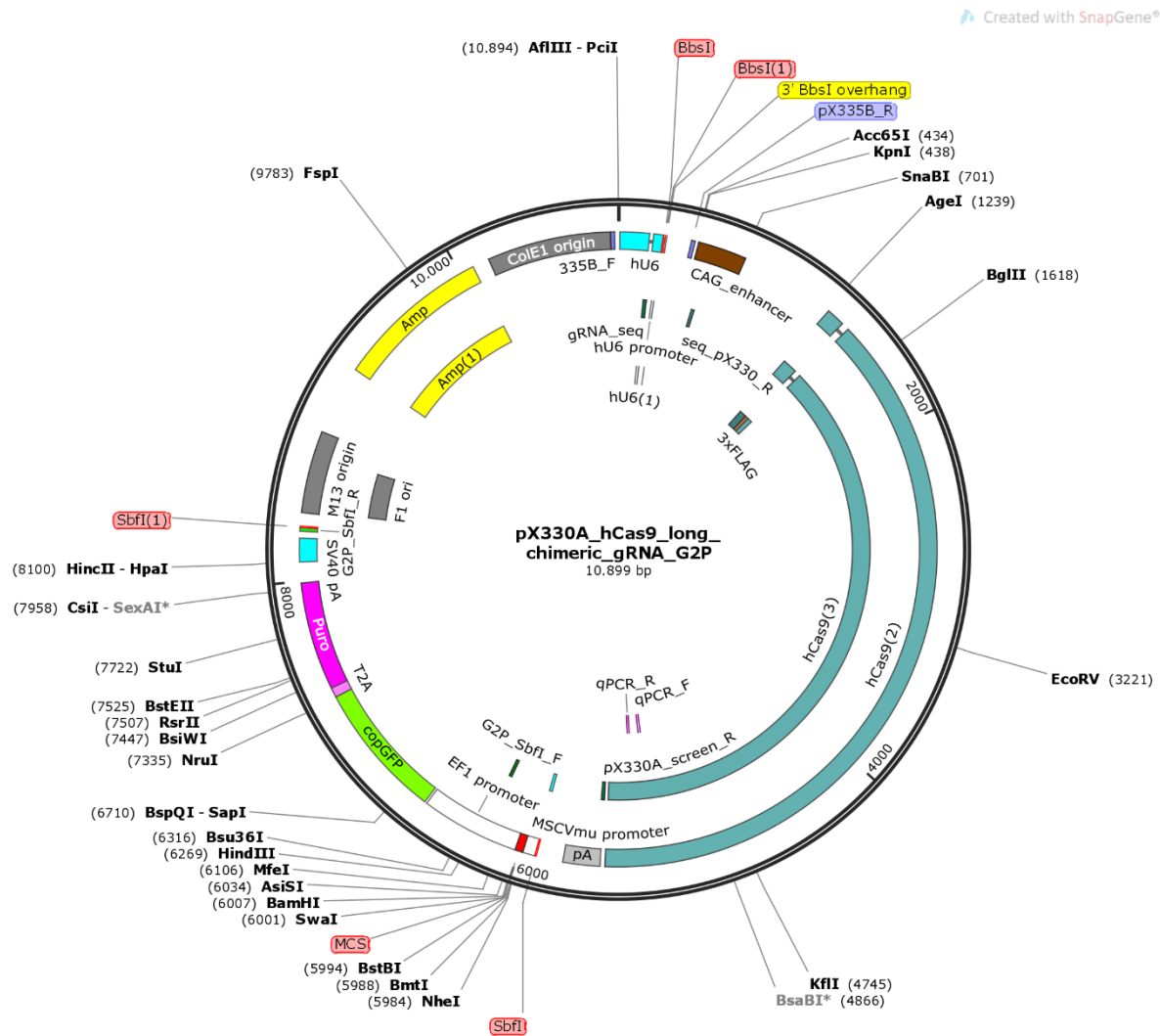


**Figure 3.1: Schematic representation of the CRISPR-Cas9 strategy used for deletion and insertion.**

#### a. gRNA design

To create a genetic deletion, two gRNAs targeting the 3' and 5' ends of the locus were designed (Figure 3.1A); for insertions, one gRNA targeting the site of insertion was designed, together with a DNA repair template containing two homology arms corresponding to the flanking regions of the insertion site (Figure 3.1B).

gRNAs were designed using Benchling CRISPR guide RNA design tool ([www.benchling.com](http://www.benchling.com)), selecting gRNAs with the overall best on-target/off-target balance scores. To clone the gRNAs into the CRISPR-Cas9 expression vector (pX330-hCas9-long-chimeric-gRNA-GFP, Figure 3.2), the gRNA sequence and the corresponding reverse complement sequence were synthesized with BbsI restriction sites overhangs.



**Figure 3.2: Scheme of pX330A\_hCas9\_long\_chimeric\_gRNA\_G2P vector.**

**Table 3.10: List of gRNAs**

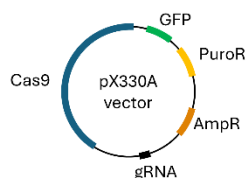
| gRNA name             | Sequence                 | Description                                            |
|-----------------------|--------------------------|--------------------------------------------------------|
| ZIC2_gRNA_del_right_F | caccGTAGAATGCAGTCACAACCG | Deletion of <i>ZIC2</i> in hiPSC                       |
| ZIC2_gRNA_del_right_R | aaacCGGTTGTGACTGCATTCTAC |                                                        |
| ZIC2_gRNA_del_left_F  | caccGGGCAGCTGAGGATTGACCT |                                                        |
| ZIC2_gRNA_del_left_R  | aaacAGGTCAATCCTCAGCTGCCC |                                                        |
| ZIC2_gRNA_ins_FHA_F   | caccATGGTACGTGTGACGGGTCG | Flag-HA tag insertion in C-ter of <i>ZIC2</i> in hiPSC |
| ZIC2_gRNA_ins_FHA_R   | aaacCGACCCGTCACACGTACCAT |                                                        |
| ZIC1_gRNA_del_right_F | caccGGGAGGGTTGGCTTAAATGT | Deletion of <i>ZIC1</i> in hiPSC                       |
| ZIC1_gRNA_del_right_R | aaacACATTTAAGCCAACCCTCCC |                                                        |
| ZIC1_gRNA_del_left_F  | caccGCCATTCATCAAGGGGGGGA |                                                        |
| ZIC1_gRNA_del_left_R  | aaacTCCCCCCTTGATGAATGGC  |                                                        |

## b. Generation of the gRNA-CRISPR-Cas9 vectors

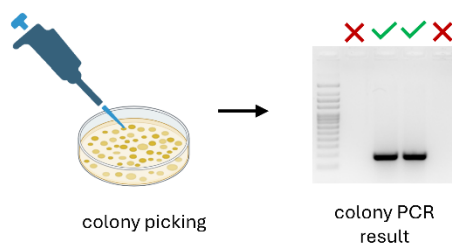
## 1 Annealing of the gRNA

5' caccGGGAGGGTTGGCTTAAATGT 3'  
3' CCCTCCAACCGAATT TACAaaa 5'

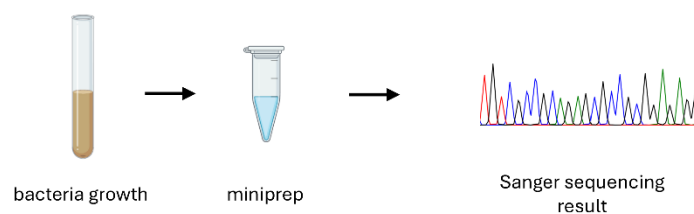
## 2 Ligation of the annealed oligos with px330A vector



## 3 Bacteria transformation &amp; positive colony selection



## 4 Miniprep preparation &amp; Sanger sequencing



**Figure 3.3: Schematic representation of the generation of the CRISPR-Cas9 expressing vector.**

1 µg of the CRISPR-Cas9 expression vector were digested with BbsI restriction enzyme at 37°C overnight. The reaction is described in Table 3.11. The digested plasmid was next column purified with the SPEEDTOOLS PCR Clean-Up kit.

Table 3.11: BbsI digestion of the CRISPR-Cas9 vector

| Component         | Amount     |
|-------------------|------------|
| Vector            | 1µg        |
| BbsI enzyme       | 1µL        |
| NEBuffer™ r2.1    | 5µL        |
| dH <sub>2</sub> O | up to 50µL |

The annealing of the oligos was performed by resuspending 1µL of each oligos at 100µM with 8µL of water followed by an incubation at 95°C for five minutes and subsequent cooling at 25°C at the rate of 5°C/min. A 1:200 dilution of the annealed oligos was ligated using T4 ligase overnight at 22°C with 50ng of the CRISPR-Cas9 expression vector, which had been previously digested with the BbsI restriction enzyme (NEB). 50µL of competent *E. coli* bacteria were transformed: bacteria were thawed on ice and mixed with 2.5µL of the ligation reaction and incubated for 1 minute on ice, followed by 1 minute at 37°C and another minute on ice; 2mL of LB medium were added to the bacteria and incubated for 1 hour at 37°C with shaking. The bacteria were centrifuged at 3500rpm for 5 minutes, and supernatant was removed letting approximately 100µL of media for the cells to be resuspended and plated on LB-ampicillin plates overnight at 37°C. The next day, colonies were picked, and colony PCR were performed to check whether the bacteria carried the intended plasmids using the forward oligo of the gRNA and a reverse primer binding to the pX330A backbone (colony\_PX330A\_R). Positive colonies were grown overnight at 37°C with shaking, in 5mL of LB media supplemented with ampicillin before plasmid purification with the NZYMiniprep kit. The selected vectors were analyzed by Sanger sequencing (using colony\_PX330A\_R primer) to confirm that the newly generated vectors contain the correct gRNA sequences. The principal steps are illustrated in Figure 3.3.

#### c. Transfection of hiPSC and clone isolation

For deletions, hiPSCs were transfected with 250ng of each gRNA vector. For the insertion, cells were transfected with 250ng of the gRNA vector and 100ng of repair template. Cells were transfected with FuGENE® HD Transfection reagent. Briefly, the DNA (i.e. gRNA vectors and repair template) and 2µL of FuGENE were mixed in 100µL of opti-MEM and incubated for 15min at room temperature before addition to the cells. The following day, the presence of green fluorescence of the transfected cells was checked before selection with puromycin during 24h to 48h. Genomic DNA from the transfected hiPSC population was extracted following instructions from

the NZY Tissue gDNA Isolation kit and PCRs were performed to confirm the presence of the genetic modification desired.

The transfected hiPSC population was seeded at single-cell density in 96 well-plates by serial dilution and grown for 10 to 15 days. Human cell media was supplemented with ROCK inhibitor (Y-27632) to promote cell survival. To evaluate the genotypes of the isolated hiPSC clones, gDNA was extracted resuspending the cells in 50µL of Quick&Dirty lysis buffer (Table 3.12) supplemented with proteinase K and incubated at 65°C for 6min followed by 2min at 98°C. The gDNA was then used for PCR using various primer combinations (Table 3.23) to distinguish between different genotypes (see Figures 4.3 and 4.4 in results section). The deletions or insertions of the selected clones were confirmed by Sanger sequencing.

#### 3.2.4. Molecular biology methods

Table 3.12: List of buffers and composition

| Buffers              | Composition                                                                                                           | Protocol                      |
|----------------------|-----------------------------------------------------------------------------------------------------------------------|-------------------------------|
| Lysis buffer 1 (LB1) | 50 mM Hepes (pH 7.5)<br>140 mM NaCl<br>1 mM EDTA<br>10% glycerol<br>0.5% NP-40<br>0.2% TX-100<br>dH2O                 | Chromatin immunoprecipitation |
| Lysis buffer 2 (LB2) | 10 mM Tris pH 8<br>200 mM NaCl<br>1 mM EDTA<br>0.5 mM EGTA<br>dH2O                                                    |                               |
| Lysis buffer 3 (LB3) | 10 mM Tris pH 8<br>100 mM NaCl<br>1 mM EDTA<br>0.5 mM EGTA<br>0.1% Na-Deoxycholate<br>0.5% N-lauroylsarcosine<br>dH2O |                               |
| RIPA ChIP buffer     | 50 mM Hepes pH 7.5<br>500 mM LiCl<br>1 mM EDTA<br>1% NP-40<br>0.7% Na-Deoxycholate<br>dH2O                            |                               |
| Elution buffer       | 50 mM Tris pH 8<br>10 mM EDTA<br>1% SDS<br>dH2O                                                                       |                               |
| Blocking solution    | PBS (1x)                                                                                                              |                               |

| Buffers                               | Composition                                                                                                               | Protocol               |
|---------------------------------------|---------------------------------------------------------------------------------------------------------------------------|------------------------|
|                                       | 0.5% BSA (w/v)                                                                                                            |                        |
| Quick&Dirty lysis buffer              | 25 mM KCl<br>5 mM Tris pH 8.3<br>1.25 mM MgCl <sub>2</sub><br>0.23% NP-40<br>0.23% Tween-20                               | Genomic DNA extraction |
| RIPA protein buffer                   | 0.2% SDS<br>1% Triton X-100<br>1mM EDTA<br>150mM NaCl<br>50 mM Tris-HCl pH 8<br>0.5% Na-Deoxycholate<br>dH <sub>2</sub> O | Protein extraction     |
| SDS-PAGE stacking gel                 | Acrylamide 30%<br>1M Tris pH 6.8<br>20% SDS<br>APS 10%<br>TEMED<br>dH <sub>2</sub> O                                      | Western blot           |
| Running gel                           | Acrylamide 30%<br>1M Tris pH 8.8<br>20% SDS<br>APS 10%<br>TEMED<br>dH <sub>2</sub> O                                      |                        |
| Blocking solution                     | 4% BSA in TBST                                                                                                            |                        |
| TBST                                  | 20 mM Tris pH 7.4<br>137 mM NaCl<br>0.05% Tween                                                                           |                        |
| Laemmli Protein Loading Buffer (5X)   | 0.5M DTT<br>10% SDS (w/v)<br>0.4 M Tris-HCl pH 6.8<br>50% Glycerol<br>8.2 mM Bromophenol Blue                             |                        |
| Running buffer                        | 25 mM Trizma Base<br>192 mM Glycine<br>0.1 % SDS                                                                          |                        |
| Transfer buffer                       | 25 mM Tris<br>190 mM Glycine<br>20% Methanol<br>0.1% SDS<br>dH <sub>2</sub> O                                             |                        |
| Enhanced Chemiluminescence Solution 1 | 0.1 mM Tris-HCl pH 8.5<br>0.4 mM Coumaric Acid<br>2.5 mM Luminol<br>dH <sub>2</sub> O                                     |                        |
| Enhanced Chemiluminescence Solution 2 | 0.1 mM Tris-HCl pH 8.5<br>0.002% Hydrogen Peroxide<br>dH <sub>2</sub> O                                                   |                        |

| Buffers      | Composition      | Protocol                   |
|--------------|------------------|----------------------------|
| Stain Buffer | 1X PBS<br>2% FBS | Single-cell RNA sequencing |

a. Protein extraction and western blot

Proteins were extracted using RIPA protein buffer supplemented with protease inhibitor. After a 30 minutes incubation on ice, the protein extract was sonicated (5 cycles at 50% amplitude, using EpiShear™ sonicator from ActiveMotif) and centrifuged at 14000rpm at 4°C for 15 minutes. Supernatant was transferred to a new tube and protein concentration measured following Pierce™ BCA Protein Assay kit's recommendation.

40µg of protein were mixed with Laemmli Buffer and denatured for 5 minutes at 95°C before being loaded on a 12% SDS-PAGE gel. Proteins were then transferred to a nitrocellulose membrane for 1 hour at 0.4A. Membrane was blocked for 1 hour at RT with blocking solution before incubation with the primary antibody at 4°C ON. The next day, the membrane was washed 3 times with TBST and incubated 1 hour at RT with the HRP-conjugated secondary antibody. Chemiluminescence substrate was applied on the membrane and the signal was captured using the Amersham ImageQuant 800 imager.

b. RNA extraction, cDNA synthesis and RNA-seq samples

Total RNA was extracted using innuPREP RNA Mini Kit according to manufacturer's instructions. 500ng of RNA were used to synthesize cDNA using ProtoScript II First Strand cDNA Synthesis Kit and oligo-dT primers. The following tables describe the reaction mix (Table 3.13) and thermocycling program (Table 3.14) used for cDNA synthesis:

Table 3.13: cDNA reaction mix

| Components                       | 20µL reaction    |
|----------------------------------|------------------|
| Template RNA                     | variable (500ng) |
| Oligo-dT                         | 2µL              |
| ProtoScript II Reaction Mix (2X) | 10µL             |
| ProtoScript II Enzyme Mix (10X)  | 2µL              |
| Nuclease-free water              | up to 20µL       |

Table 3.14: cDNA synthesis

| Temperature | Time  |
|-------------|-------|
| 25°C        | 5min  |
| 42°C        | 1hour |
| 80°C        | 5min  |

For RNA-seq experiments, at least 1µg of the RNA samples were shipped to Macrogen Inc. (Korea), where the samples were treated with DNase before library preparation. Libraries were prepared following the TruSeq stranded mRNA library (Illumina) protocol and sequenced as paired end 150bp (<40M reads/sample) in a NovaSeq6000 sequencer.

#### c. Single-cell RNA-seq

On day 8 of the differentiation of hiPSC into anterior neural progenitors, cells were detached from the plates with TrypLE™ Express for 15 to 20 minutes at 37°C. One million cells of each condition were resuspended in 180µL of stain buffer and labelled with 20µL of Sample Tags during 20 minutes at room temperature (WT untreated: sample tag1, WT treated with 0.5µM of SAG: sample tag 2, *ZIC2*<sup>-/-</sup> untreated: sample tag3, *ZIC2*<sup>-/-</sup> treated with 0.5µM of SAG: sample tag4) to allow multiplexing. After incubation, 2mL of stain buffer were added before centrifuging at 500g for 5 minutes at room temperature. The samples were next resuspended in 200µL of cold sample buffer (from the BD Rhapsody™ Cartridge Reagent Kit) and 7500 cells of each sample were counted before merging them into 620µL of sample buffer. Cells were then subject to Single Cell RNA capture. Single Cell RNA capture, reverse transcription and libraries preparation were performed as described in the BD Rhapsody Single-Cell Analysis protocol. Libraries were then sequenced at a depth of 40000 reads using 150bp paired end reads. These experiments have been performed in collaboration with Nacho Varela's laboratory (IBBTEC, Santander, Spain).

#### d. Chromatin immunoprecipitation and ChIP-seq

A confluent 10cm plate was cross-linked with 1% formaldehyde for 10 minutes at RT and quenched with 0.125M of glycine for another 10 minutes. Cells were washed two times with cold PBS supplemented with proteinase inhibitor and harvested from the plate and transferred to a 15mL falcon tube. Cross-linked cells were centrifuged at 1200rcf, for 5 minutes at 4°C and the supernatant was removed.

Next, the cross-linked cells were sequentially resuspended in three lysis buffers: 5mL of Lysis Buffer 1 (LB1) for 10 minutes at 4°C rotating, 5mL of Lysis Buffer 2 (LB2) for 10 minutes at RT

rotating and 1 mL of Lysis Buffer 3 (LB3). The isolated chromatin was maintained on ice for the rest of the protocol. Chromatin was sonicated for 19 cycles at 30% amplitude, using EpiShear™ sonicator from ActiveMotif.

5% of the sonicated chromatin was saved as the input sample and ~75% of the remaining chromatin was used for transcription factors ChIP and ~25% for histone marks. Volume of the chromatin was adjusted to 900µL adding LB3 and completed to 1mL by adding 100µL of Triton. Antibody (3.5µg for histone modification and 8µg for transcription factor) was bound to the chromatin overnight at 4°C rotating.

50µL of Protein-A magnetic beads blocked with a 0.5% BSA solution were incubated for 4 hours at 4°C rotating with the sonicated chromatin for the antibody to bind the beads. Beads were next washed 5 times with RIPA Buffer and chromatin was eluted in 210µL of Elution Buffer at 65°C for 15 minutes with shaking. Chromatin was de-crosslinked together with the input samples at 65°C overnight and treated with RNase A for 1 hour at 37°C followed by a proteinase K treatment at 55°C for 2 hours. DNA was purified using the QIAquick PCR & Gel Cleanup Kit.

The input and ChIP DNAs were then shipped to Macrogen Inc, in order to generate ChIP-seq libraries using the Illumina TruSeq ChIP-seq kit. The resulting libraries were sequenced at depth of 40M reads using 150bp paired end reads.

#### e. Polymerase Chain Reaction

Table 3.15: Q5 reaction mix

| Components            | 25µL reaction |
|-----------------------|---------------|
| 5X Q5 reaction buffer | 5µL           |
| 10mM dNTPS            | 0.5µL         |
| 10µM Forward Primer   | 1.25µL        |
| 10µM Reverse Primer   | 1.25µL        |
| Q5 polymerase         | 0.25µL        |
| DNA template          | variable      |
| Nuclease-free water   | up to 25µL    |

Table 3.16: Q5 Thermocycling program

| Steps                | Temperature | Time   |
|----------------------|-------------|--------|
| Initial denaturation | 98°C        | 30s    |
| 35-40 cycles         | 98°C        | 10s    |
|                      | 56°C – 65°C | 15s    |
|                      | 72°C        | 20s/kb |
| Final extension      | 72°C        | 2min   |

Table 3.17: OneTaq reaction mix

| Components          | 25µL reaction |
|---------------------|---------------|
| 5X OneTaq GC Buffer | 5µL           |
| 10mM dNTPS          | 0.5µL         |
| 10µM Forward Primer | 0.5µL         |
| 10µM Reverse Primer | 0.5µL         |
| OneTaq polymerase   | 0.125µL       |
| DNA template        | variable      |
| Nuclease-free water | up to 25µL    |

Table 3.18: OneTaq Thermocycling program

| Steps                | Temperature | Time    |
|----------------------|-------------|---------|
| Initial denaturation | 94°C        | 30s     |
| 35-40 cycles         | 94°C        | 15s     |
|                      | 56°C – 65°C | 30s     |
|                      | 68°C        | 1min/kb |
| Final extension      | 68°C        | 5min    |

Table 3.19: NZYTaq II reaction mix

| Components          | 25µL reaction |
|---------------------|---------------|
| NZYTaq II 2×        | 12.5µL        |
| 10µM Forward Primer | 0.5µL         |
| 10µM Reverse Primer | 0.5µL         |
| DNA template        | variable      |
| Nuclease-free water | up to 25µL    |

Table 3.20: NZYTaq II Thermocycling program

| Steps                | Temperature | Time   |
|----------------------|-------------|--------|
| Initial denaturation | 95°C        | 3min   |
| 35-40 cycles         | 94°C        | 30s    |
|                      | 56°C – 65°C | 30s    |
|                      | 72°C        | 30s/kb |
| Final extension      | 72°C        | 5min   |

Table 3.21: Quantitative PCR reaction

| Components                      | 10µL reaction |
|---------------------------------|---------------|
| NZYSpeedy qPCR Green Master Mix | 5µL           |
| DNA                             | 0.4µL         |
| 10µM F and R primers            | 0.125µL       |
| Nuclease-free water             | up to 10µL    |

Table 3.22: Quantitative PCR program

| Steps                | Temperature  | Time |
|----------------------|--------------|------|
| Initial denaturation | 95°C         | 2min |
| 40 cycles            | 95°C         | 5s   |
|                      | 60°C         | 30s  |
| Melting Curve        | 95°C         | 15s  |
|                      | 60°C         | 1min |
|                      | 60°C to 95°C | -    |

Table 3.23: List of primers

| Primer name     | Sequence               | Description                                                                |
|-----------------|------------------------|----------------------------------------------------------------------------|
| colony_PX330A_R | GGAAAGTCCCTATTGGCGTT   | Binds to pX330A vector backbone                                            |
| ZIC2_F1         | CATCTGGGGAAATTCGTGGC   | Identify clones with <i>ZIC2</i> deletion                                  |
| ZIC2_R2         | CACCTTCCCTTGACCATCCT   |                                                                            |
| ZIC2_F3         | AAAATGAGCCGTGCCAAAGT   |                                                                            |
| ZIC2_R4         | AGTTCACGGTCCTGCATCTC   |                                                                            |
| ZIC2_F6         | GGCTTTACTGTGGTTTCGCA   |                                                                            |
| ZIC2_R7         | TTGCTACGTGTTGTTGGGG    |                                                                            |
| ZIC2_SNP_F      | AGCCAGAAAATTAAACGGGGAG | Confirm the presence of only one allele in the heterozygous clone isolated |
| ZIC2_SNP_R      | CAGCCCTCAAACCTCACACTG  |                                                                            |
| ZIC2_FHA_F1     | GGCCTCTCCTCCAACTTCAA   | Identify the insertion of the repair template                              |
| ZIC2_FHA_R2     | TCAGCTTCAAAGACTCCGGA   |                                                                            |
| ZIC1_F1         | TACCTGGGATTGATGAGGCG   | Identify clones with <i>ZIC1</i> deletion                                  |
| ZIC1_R2         | GCTACACACAGGAAACAGCT   |                                                                            |
| ZIC1_R6         | GCGGTTTATCTTCCTGGGGA   |                                                                            |
| ACTB_exp_F      | CACCCAGCACAATGAAGATCA  | Measure the expression of the given gene by RT-qPCR                        |
| ACTB_exp_R      | CCTGCTTGCTGATCCACATC   |                                                                            |
| CTSF_exp_R      | GCTGGCCATGGTGTTAC      |                                                                            |
| CTSF_exp_F      | TTTGGGCCATCAAGAACAGC   |                                                                            |
| EEF2_exp_F      | CTATCTGCCCCGTCAACGAGT  |                                                                            |
| EEF2_exp_R      | GATCTGCCAGTGGTCAAACA   |                                                                            |
| HES5_exp_F      | AAGCTGGAGAAGGCCGACAT   |                                                                            |
| HES5_exp_R      | CCTTCGCTGTAGTCCTGGTG   |                                                                            |
| LMX1A_exp_F     | CATCGAGCAGAGTGTCTACAGC |                                                                            |
| LMX1A_exp_R     | TGTCGTCGCTATCCAGGTCATG |                                                                            |
| PAX6_exp_F      | GCCAGACCTCCTCATACTCC   |                                                                            |
| PAX6_exp_R      | TGACACACCAGGGGAAATGA   |                                                                            |
| ZIC1_exp_F      | CGACAAGTCCTACACGCATC   |                                                                            |
| ZIC1_exp_R      | AATTGGAAGAGAGCGCACTG   |                                                                            |
| ZIC2_exp_F      | GATGTGCGACAAGTCCTACAC  |                                                                            |

| Primer name           | Sequence              | Description                |
|-----------------------|-----------------------|----------------------------|
| ZIC2_exp_R            | TGGACGACTCATAGCCGGA   |                            |
| chr4_neg_F            | GAACTCCCAGACCGACAGAA  | Negative control ChIP-qPCR |
| chr4_neg_R            | TTCCCACATGTCCCCATTCC  |                            |
| chr8_neg_F            | GCGCCTCAACATGACTTTGA  |                            |
| chr8_neg_R            | TGTGAAGAGGGTCCAGTCTG  |                            |
| LMX1B_pos_ChIP_ZIC2_F | GGCAGAGACCTTTCAGACCT  |                            |
| LMX1B_pos_ChIP_ZIC2_R | TAATCGCCCCGCAGTCATTTG | Positive control ChIP-qPCR |

### *Colony PCR*

To confirm whether the plasmid of interest was present in the bacterial colonies picked during cloning, colony PCRs were performed: the bacterial colonies were resuspended in 20µL of water and 3µL were incubated at 95°C for 5min before use within the PCR reaction using NZYtaq II 2x Green Master Mix. PCR reaction and thermocycling program are described in Table 3.19 and 3.20, respectively.

### *Genotyping PCR*

To genotype the new cell lines generated with CRISPR-Cas9 technology, PCR were performed using OneTaq polymerase or NZYtaq II 2x Green Master Mix depending on the locus modified. PCR reactions were prepared as described in Polymerase Chain Reaction section.

### *Agarose gel electrophoresis*

In order to visualize PCR products, samples were mixed with orange G dye and loaded on 1-2% agarose gel (prepared in 1X TAE buffer and GreenSafe Premium), depending on the expected size and run at 100V for 30 minutes. Size of PCR products was determined using NZYDNA Ladder VII.

Pictures of the gel were taken with the Chemidoc XRS transilluminator from Biorad. When necessary, bands were cut out of the gel and DNA was purified using SpeedTools PCR Clean-up kit.

### *Quantitative PCR*

To measure gene expression, quantitative PCRs were performed using NZYSpeedy qPCR Green Master Mix as described in Tables 3.21 and 3.22, using the CFX Opus 384 Real-Time PCR System from BioRad. Prior to the reaction preparation, cDNA samples were diluted 1:4. Forward and reverse primers were designed in a different exon, to avoid amplifying genomic DNA.

## 3.2.5. Data analysis

Table 3.24: List of software

| Software                                                              | Reference                                                                                                                           |
|-----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| BD Rhapsody™ Sequence Analysis Pipeline                               | <a href="https://bd-rhapsody-bioinfo-docs.genomics.bd.com/">https://bd-rhapsody-bioinfo-docs.genomics.bd.com/</a>                   |
| bedtools                                                              | (Quinlan & Hall, 2010)                                                                                                              |
| Benchling                                                             | <a href="http://www.benchling.com">www.benchling.com</a>                                                                            |
| Bowtie2                                                               | (Langmead & Salzberg, 2012)                                                                                                         |
| clusterProfiler                                                       | (Wu et al., 2021) version 4.4.4                                                                                                     |
| Cufflinks                                                             | (Trapnell et al., 2010)                                                                                                             |
| Deeptools<br>(bamCoverage, bigwigaverage, computematrix, plotheatmap) | (Ramírez et al., 2016)                                                                                                              |
| DEseq2                                                                | (Love et al., 2014) version 1.36.0                                                                                                  |
| Diffbind                                                              | (Stark & Brown, n.d.) version 3.6.5                                                                                                 |
| edgeR                                                                 | (Chen et al., 2025) version version 3.38.4                                                                                          |
| effsize                                                               | <a href="https://CRAN.R-project.org/package=effsize">https://CRAN.R-project.org/package=effsize</a> version 0.8.1                   |
| FactoMineR                                                            | (Lê et al., 2008) version 2.11                                                                                                      |
| fastqc                                                                | <a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a> |
| Figeno                                                                | (Sollier et al., 2024) version 1.4.0                                                                                                |
| GREAT                                                                 | (McLean et al., 2010) version 4.0.4                                                                                                 |
| Hisat2                                                                | (D. Kim et al., 2019)                                                                                                               |
| Htseq                                                                 | (Anders et al., 2015)                                                                                                               |
| Macs2                                                                 | (Feng et al., 2012)                                                                                                                 |
| MEME-CHIP                                                             | (Machanick & Bailey, 2011) version 5.5.4                                                                                            |
| pheatmap                                                              | <a href="https://cran.r-project.org/package=pheatmap">https://cran.r-project.org/package=pheatmap</a> version 1.0.12                |
| samtools                                                              | (H. Li et al., 2009)                                                                                                                |
| Seurat                                                                | (Hao et al., 2024) version 5.1.0                                                                                                    |
| trimmomatic                                                           | (Bolger et al., 2014)                                                                                                               |

### a. qPCR analysis

Relative gene expression levels were calculated with the  $2^{\Delta Ct}$  method using *EEF2* and *ACTB* as housekeeping genes. Standard deviations were represented as error bars and calculated from technical triplicates for each sample.

Enrichment of ChIP-qPCR experiments was calculated as percentage input. Technical triplicates were used for both, samples and input. First, the Ct of the 1% input was adjusted to 100% as indicated:  $adj\_input = Ct\_input - \log_2(100)$ . Second, the enrichment was calculated as follows:  $100 \times 2^{(adj\_input - Ct)}$ . Standard deviations of the technical triplicates were calculated and represented as error bars.

### b. RNA-seq analysis

Computational analysis of the RNA-seq data was performed as follows. First, quality control was performed, trimming out sequencing adaptors and filtering out low quality reads using trimmomatic (Bolger et al., 2014) and fastqc (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Next, filtered reads were mapped on the reference genome hg19 with hisat2 (D. Kim et al., 2019). Only the reads with a mapping quality greater than 10 and properly mapped were kept with SAMtools (H. Li et al., 2009). FPKMs, counts and bigwig files were next generated using cufflinks (Trapnell et al., 2010), htseq (Anders et al., 2015) and bamCoverage (Ramírez et al., 2016), respectively.

Principal component analyses (PCA) were performed with the PCA function from the FactoMineR package (Lê et al., 2008).

Prior to differential expression analysis, counts per million (cpm) were calculated using the cpm function from edgeR package (Y. Chen et al., 2025). Differentially expressed genes were identified using the DEseq function from DEseq2 package (Love et al., 2014). To be considered differentially expressed, a gene has to respect the following criteria: *FPKM* > 1 in at least one condition,  $|\log_2(\text{FoldChange})| > 1$  and *p-adj* < 0.01. Heatmaps of the differentially expressed genes were plotted using pheatmap function (<https://cran.r-project.org/package=pheatmap>).

Gene Ontology (GO term analyses) of biological processes were conducted with the enrichGO function from clusterProfiler R package (Wu et al., 2021). The default cutoffs were used (*p*-value < 0.05, *q*-value < 0.2). To avoid redundancy of the GO terms, the simplify function from clusterProfiler with the default parameters was used. The 10 first GO term results are represented as barplots, ordering them by  $-\log_{10}(p\text{-adj})$ .

### c. ChIP-seq analysis

ChIP-seq experiments were performed in duplicate for each condition. First, quality control was performed, trimming out sequencing adaptors and filtering out low quality reads using trimmomatic and fastqc, respectively. Second, reads were mapped on hg19 genome using bowtie2 (Langmead & Salzberg, 2012), generating bam files. Only the reads with a mapping quality greater than 10 and properly mapped were kept with SAMtools (H. Li et al., 2009), duplicated reads were also filtered out. Bigwig files were generated using bamCoverage (Ramírez et al., 2016). Using the bam files generated, peak calling was performed using macs2 (Feng et al., 2012), with a q-value cutoff of 0.05. For histone marks (H3K27ac and H3K27me3 ChIP), the broad peak option is specified. The narrow peak option was used for ZIC2 ChIP. The resulting bed files were handled using the bedtools suite (Quinlan & Hall, 2010). The bed file of the coordinates of TSS were downloaded from biomaRt (<https://www.ensembl.org/index.html>) and coordinates of CpG islands (CGI) were obtained from Illingworth et al., 2010 and Long et al., 2013.

To identify the differential binding sites, we used the DiffBind package in R (Stark & Brown, n.d.). A DBA object was created using the dba function, which takes as inputs the bam file of each sample and a bed file containing the union of the peaks from the samples being compared. The following DiffBind functions were used with the default parameters: dba.count to count reads in binding site intervals, dba.contrast to define the groups to be compared, dba.analyze to perform differential peak analysis using DESeq2 package, and dba.report to extract the differentially bound sites, applying a false discovery rate (FDR) cutoff of 0.05.

Motif discovery analyses were performed with MEME-CHIP (Machanick & Bailey, 2011) in classic mode. As input, the analysis used motifs from the Hocomoco database (human and mouse orthologs).

Regions of interest (i.e., regions bound by ZIC2 that are either gaining or losing H3K27ac) were associated with putative target genes using GREAT (McLean et al., 2010) with the basal plus extension mode. In this mode, each gene is assigned a basal regulatory domain (5kb upstream and 1kb downstream of the TSS) which is extended towards the nearest gene's basal domain, with a maximum extension of 1000kb.

For visualization purposes, the bigwigs of each duplicate were averaged with bigwigaverage function (Ramírez et al., 2016). Heatmaps were generated with the computematrix (that calculates scores from bigwig files for the specified regions within a bed file) and plotheatmap (that plot the heatmap of the previously calculated scores) functions from the deeptools suite (Ramírez et al., 2016). Bigwig tracks figures were generated using Figeno tool (Sollier et al., 2024).

#### d. Single-cell RNA-seq analysis

The initial analysis of the single-cell RNA-seq data was performed using BD Rhapsody™ Sequence Analysis Pipeline (<https://bd-rhapsody-bioinfo-docs.genomics.bd.com/>). In summary, the pipeline takes as input the fastq files, performs quality filtering, identifies the cell barcode and unique molecular identifier (UMI) and aligns the reads to the human reference genome (hg38). Next, raw reads counts are calculated, and error correction is applied, and it associates the reads with their corresponding cells. Sample Tag Analysis is next performed to associate a putative cell with its sample of origin. Next, the expression matrix table is generated and preloaded into Seurat package format. This analysis pipeline was performed by Natalia Sánchez Collantes, from Nacho Varela laboratory (IBBTEC, Santander, Spain).

The resulting Seurat object was handled with Seurat toolkit (Hao et al., 2024) in R. First, quality filtering of the data was applied, keeping only the cells with unique feature counts between 2500 and 8000, and cells expressing less than 20% of mitochondrial genes. Low quality cells will often have few genes or on the other hand, a too high gene count might be related to doublets or multiplets. Similarly, a high percentage of mitochondrial genes is associated with dying cells. Next, the data were normalized with the default parameter of the `NormalizeData` function. Differentially expressed genes between samples were identified using the `FindMarkers` function. In order to visualize the data, we run the UMAP dimensional reduction and plotted it thanks to the `Featureplot`, `Featureplot3` and `Dimplot` functions. The dotplot were generated using the `Dotplot` function.

#### e. Integration of ChIP-seq and RNA-seq data

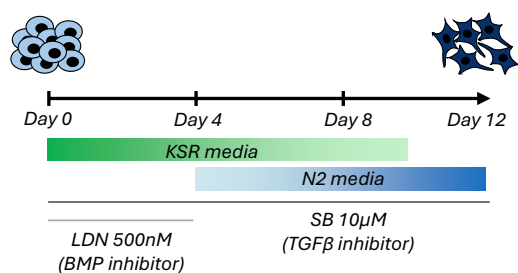
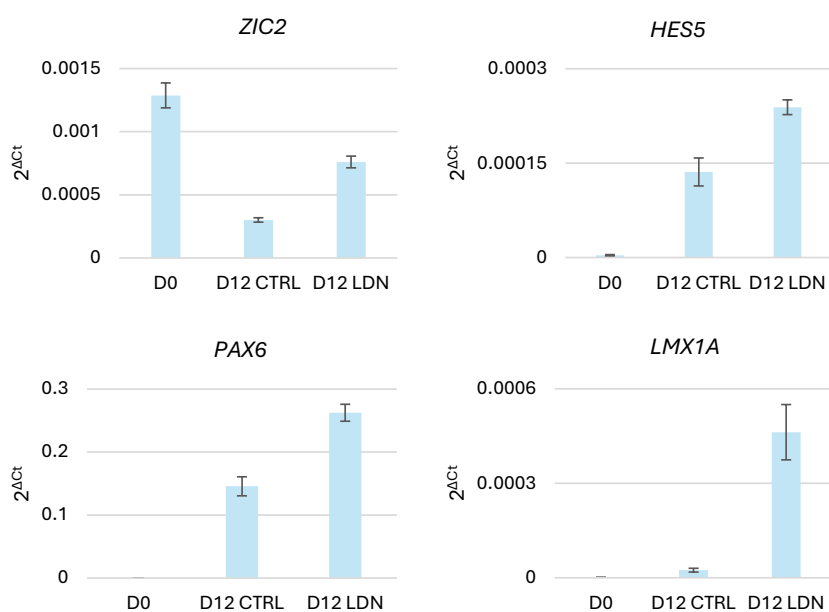
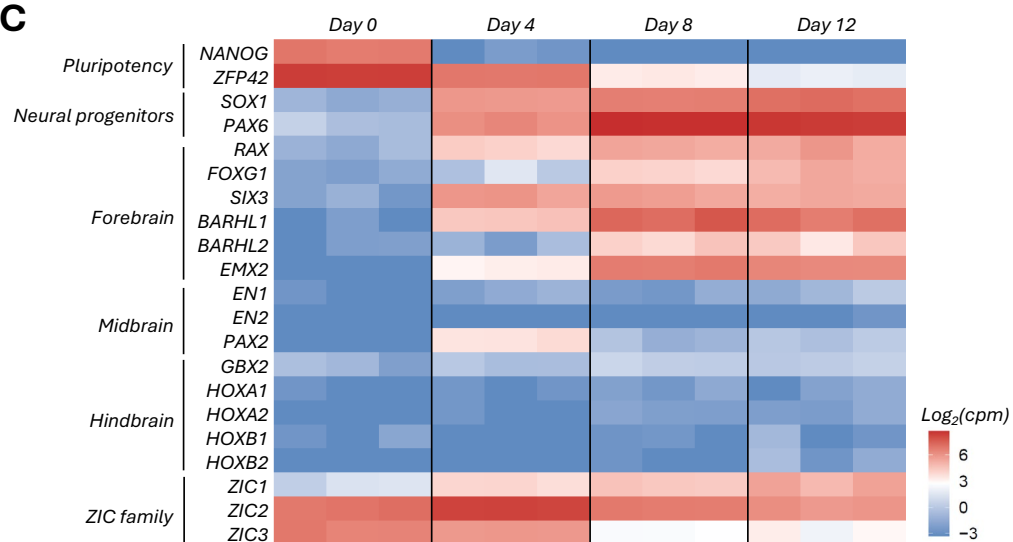
In order to test whether the gain or loss of acetylation at ZIC2-bound regions correlates with changes in expression, we calculated the distance of each gene to the closest ZIC2-bound regions gaining or losing H3K27ac with `bedtools closest` function (Quinlan & Hall, 2010). Next, the cumulative distribution of the distance was plotted, separating the genes into three groups: up-regulated genes, down-regulated genes and all genes (hg19). To test whether the distribution of two groups was significantly different, we performed Wilcoxon statistical tests. The magnitude of the effect size (quantification of the difference between two groups) was assessed by computing the Cliff's Delta effect size, using `cliff.delta` function from `effsize` package in R (<https://CRAN.R-project.org/package=effsize>). A cliff delta value lower than 0.147 is considered negligible, lower than 0.33 is small, lower than 0.474 is medium and otherwise is considered a large effect size.

## 4. Results

#### 4.1. Differentiation of WT hiPSC into anterior neural progenitor cells (AntNPC)

Previous clinical and *in vivo* studies indicate that *ZIC2* plays a relevant role in the establishment of the dorsal-ventral axis during early forebrain development and particularly in the establishment of dorsal identities during neurulation. However, the molecular mechanisms (e.g. *ZIC2*-dependent regulatory networks) whereby *ZIC2* executes these important developmental functions remain unclear, particularly in humans. In this thesis, we took advantage of an *in vitro* differentiation protocol whereby hiPSC are differentiated into AntNPC to model early brain development and explore the role of *ZIC2* during the induction of anterior neural identities (Figure 4.1A). This differentiation protocol is based on dual SMAD-inhibition, using SB and LDN molecules, that inhibit TGF $\beta$  and BMP signaling pathways, respectively. To favor dorsal identities, the addition of LDN to the media was limited to the first four days of differentiation. By RT-qPCR, we confirmed the up-regulation of *LMX1A*, *PAX6* and *ZIC2* genes marking dorsal identities, compared to the original protocol (Figure 4.1B). Expression of the neural marker *HES5* was induced in both conditions (Figure 4.1B). This modified AntNPC differentiation protocol (i.e. LDN treatment during the first four days only) was used in all subsequent experiments.

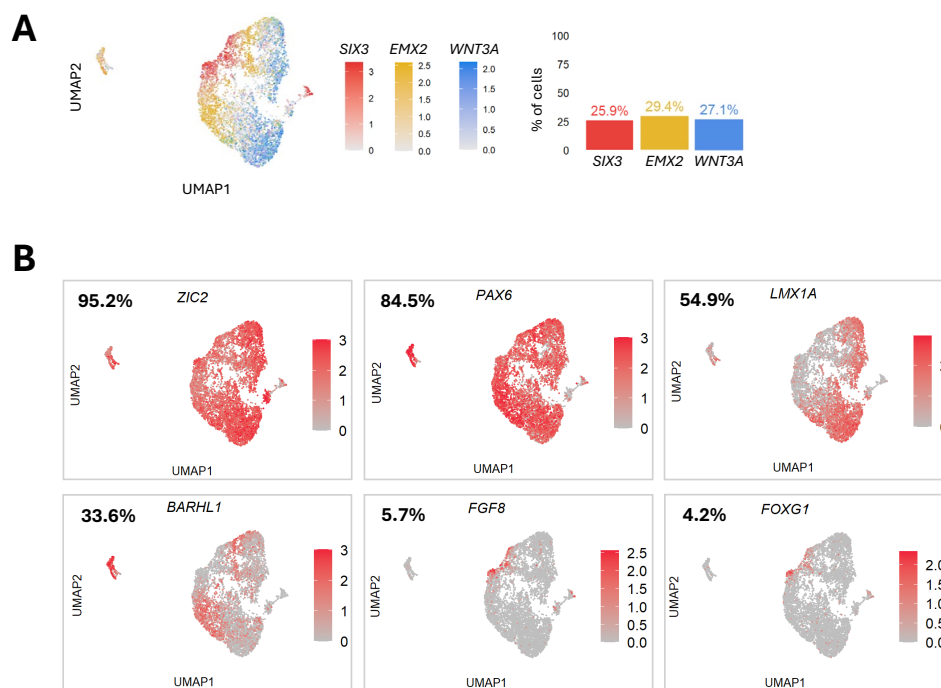
WT hiPSC were differentiated and RNA was collected at different timepoints of the AntNPC differentiation (day 0, day 4, day 8 and day 12). The resulting RNAs were then analyzed by RNA-seq in order to more globally investigate the gene expression dynamics and the cellular identities obtained with this differentiation system (Figure 4.1C). We confirmed the downregulation of pluripotency marker genes (*NANOG*, *ZFP42*) and the upregulation of neural progenitor (*SOX1*, *PAX6*) and forebrain identity genes (*RAX*, *FOXG1*, *SIX3*, *BARHL1*, *BARHL2*, *EMX2*) after differentiation. The induction of many of these neural and forebrain markers already started on day 4 and peaked by day 8. Markers of more posterior identities of the central nervous system, namely, the midbrain (*EN1*, *EN2*, *PAX2*) and hindbrain (*GBX2*, *HOXA1*, *HOXA2*, *HOXB1*, *HOXB2*) were not induced. On the other hand, *ZIC2* was already expressed on day 0 and its expression was maintained up to day 12 with a higher expression on day 4. Regarding other *ZIC* family members, *ZIC1* was stably expressed from day 4 to day 8 and showed the highest expression on day 12, while *ZIC3* expression was highest on day 0 and day 4 and then decreased from day 8 onwards.

**A****B****C**

**Figure 4.1: Differentiation of WT hiPSC into AntNPCs.**

(A) Schematic representation of the 12-day differentiation protocol adapted from Tchieu et al., 2017. (B) Expression of key marker genes was measured by RT-qPCR in WT cells. D12 CTRL sample corresponds to the differentiated WT cells on day 12, following the original protocol. D12 LDN sample corresponds to the differentiated WT cells on day 12, limiting LDN addition to the first four days of the differentiation. Expression levels were calculated using the  $2^{\Delta Ct}$  method and the standard variation of technical triplicates is represented as error bars. (C) Heatmap showing the expression dynamics (as measured by RNA-seq) of genes considered as markers of pluripotency, neural progenitors and different brain regions along the rostro-caudal axis. The expression dynamics of several ZIC family genes is also shown. The color scale is shown as  $\log_2(\text{counts per million})$ .

To further characterize the cellular identities obtained with our modified AntNPC differentiation protocol, we also collected WT cells on day 8 to perform single-cell RNA-seq (Figure 4.2). This experiment revealed that the differentiation protocol recapitulates important aspects of dorsal-ventral and rostro-caudal forebrain patterning. Indeed, when plotting the expression of key neural markers in a UMAP space, we can easily distinguish different expression domains that seem to represent distinct positional identities (Figure 4.2A). Cells expressing dorsal telencephalon markers like *WNT3A* or *LMX1A* are on the right side of the UMAP plot while markers of the posterior telencephalon/diencephalon, such as *EMX2* or *BARHL1*, are localized in the middle of the UMAP. Finally, markers of the rostro-ventral telencephalon, namely, *FGF8*, *FOXP1* or *SIX3* are expressed in the cells located on the left side of the UMAP (Figure 4.2A and 4.2B). This *in vitro* differentiation model of hiPSC into AntNPC should therefore provide a useful framework to investigate the role of *ZIC2* during neural induction and patterning of anterior neural progenitors.

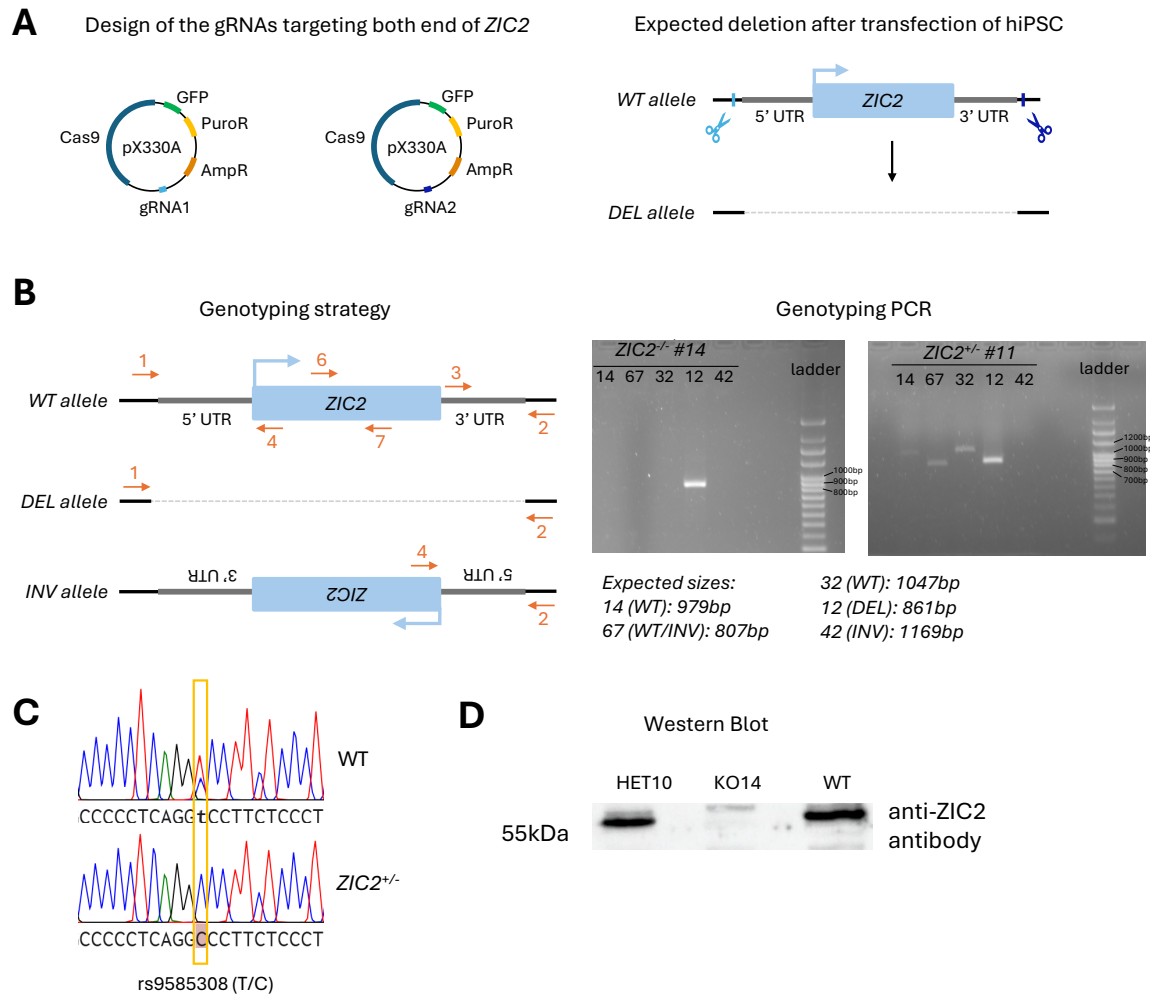


**Figure 4.2: Differentiation of WT hiPSC into AntNPCs partially recapitulates forebrain patterning.**

(A) UMAP plot showing the expression of key markers of different forebrain regions in WT AntNPC on day 8 (scRNA-seq data). Expression of SIX3 is shown in red, EMX2 in yellow and WNT3A in blue. (B) UMAP plots showing the expression of selected genes within day 8 AntNPC. The percentage of cells expressing each marker is shown in the upper-left corner. Expression is shown as  $\log_{10}((\text{feature counts per cell} / \text{total counts per cell}) \times 10000)$ .

#### 4.2. Genetic engineering of hiPSC

To interrogate *ZIC2* regulatory function and the molecular defects associated with its loss during the AntNPC differentiation, a cell line in which *ZIC2* is absent was required. To do so, we used CRISPR-Cas9 technology to delete the *ZIC2* gene. Briefly, gRNAs targeting both ends of the gene were designed and cloned within the pX330A vector expressing the Cas9 nuclease (Figure 4.3A). Then, hiPSC were transfected with the resulting vector in order to introduce double-strand breaks that, upon repair by non-homologous end joining (NHEJ), could lead to the deletion or inversion of the *ZIC2* gene. By designing primers binding along the *ZIC2* locus, we were able to distinguish between the different alleles produced by the CRISPR editing approach (Figure 4.3B). We successfully isolated several hiPSC clones carrying the *ZIC2* deletion, either in a homozygous (#14, #25 and #34) or in a heterozygous manner (#10, #11 and #28) (Figure 4.3B). To confirm that the identified heterozygous cell lines were truly heterozygous for the deletion, we took advantage of a SNP (rs9585308) located within the *ZIC2* gene and that was heterozygous in the parental WT hiPSC line. Sanger sequencing confirmed that in the *ZIC2*<sup>+/-</sup> lines, only one SNP allele was present (Figure 4.3C). In addition, we confirmed by western blot, the presence (WT and *ZIC2*<sup>+/-</sup> lines) or absence (*ZIC2*<sup>-/-</sup> lines) of the *ZIC2* protein (Figure 4.3D).

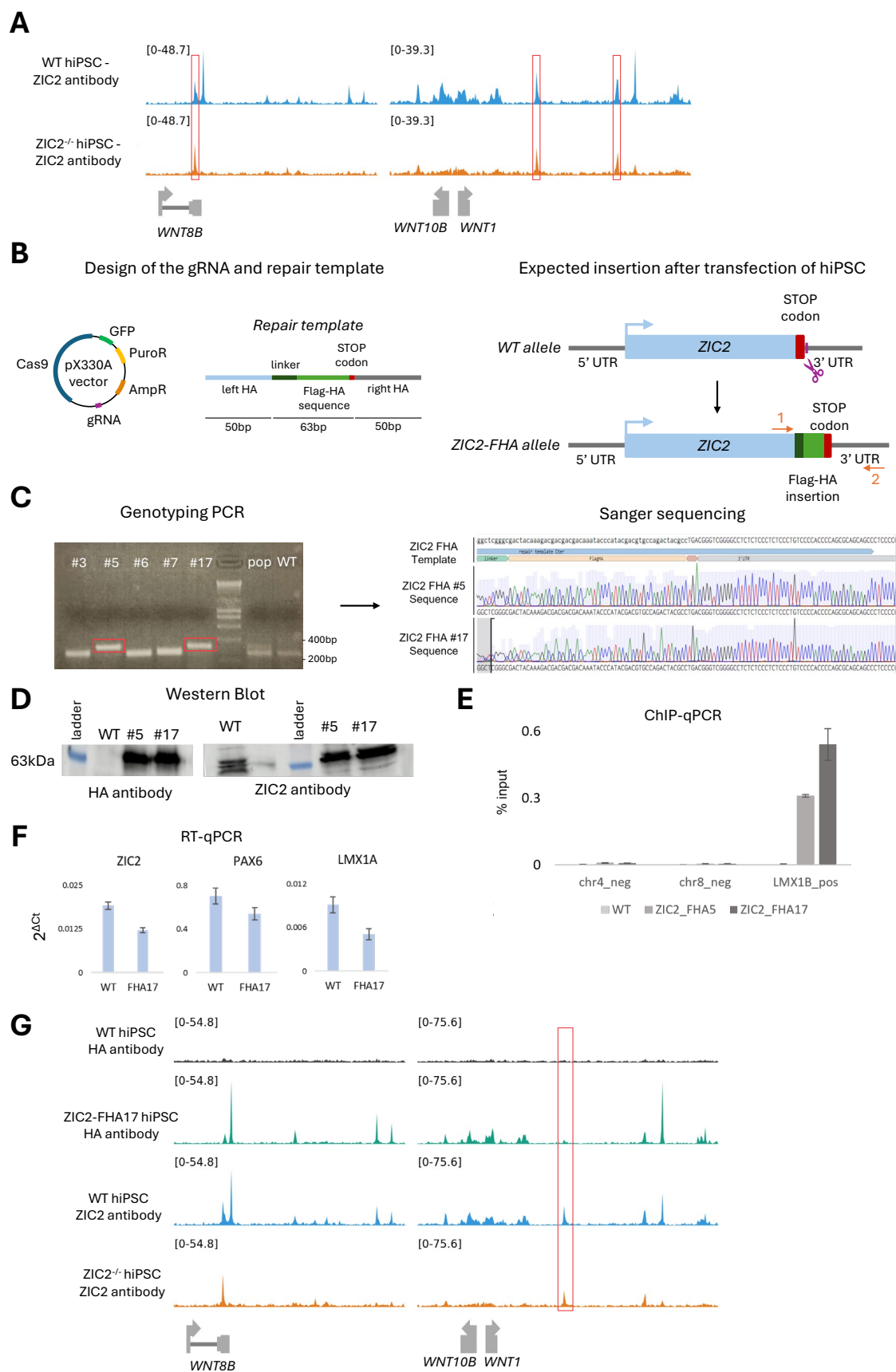


**Figure 4.3: Generation of *ZIC2*<sup>-/-</sup> and *ZIC2*<sup>+/-</sup> hiPSC lines with CRISPR-Cas9.**

(A) Schematic representation of the CRISPR-Cas9 strategy used to delete the *ZIC2* gene. Each pX330A vector expresses the Cas9 nuclease together with a gRNA targeting one end of the *ZIC2* gene (blue scissors). (B) Genotyping strategy used to confirm the heterozygous or homozygous deletion of *ZIC2* in the isolated clones. The different combinations of the primers (orange arrows) allow for the discrimination between three possible alleles (WT, deletion (DEL) or inversion (INV)) generated by the CRISPR cuts. Several homozygous (#14, #25 and #34) and heterozygous (#10, #11 and #28) hiPSC lines for the *ZIC2* deletion were isolated. Examples of genotyping PCRs are shown for the clones #11 (heterozygous) and #14 (homozygous). (C) Sanger sequencing result of the rs9585308 SNP in WT and *ZIC2*<sup>+/-</sup> hiPSC. (D) Western blot analysis of WT, *ZIC2*<sup>+/-</sup> and *ZIC2*<sup>-/-</sup> hiPSC, using the *ZIC2* antibody confirmed the presence (WT and *ZIC2*<sup>+/-</sup> lines) or absence (*ZIC2*<sup>-/-</sup> line) of the *ZIC2* protein.

Although a commercial ChIP-grade antibody targeting *ZIC2* is available (ab150404, abcam), our data suggests that it is not exclusively recognizing *ZIC2*, most likely due to cross-reactivity with other *ZIC* family members because of their high homology (Houtmeyers et al., 2013). Indeed, we performed a ChIP-seq experiment in WT and *ZIC2*<sup>-/-</sup> hiPSC, using the commercial *ZIC2* antibody. In *ZIC2*<sup>-/-</sup> cells, some peaks could still be detected by the antibody, indicating that it is not fully specific for *ZIC2* (Figure 4.4A). To overcome this limitation, we decided to tag *ZIC2* with a Flag-HA epitope using CRISPR-Cas9 technology (Figure 4.4B). Briefly, by generating a double-strands

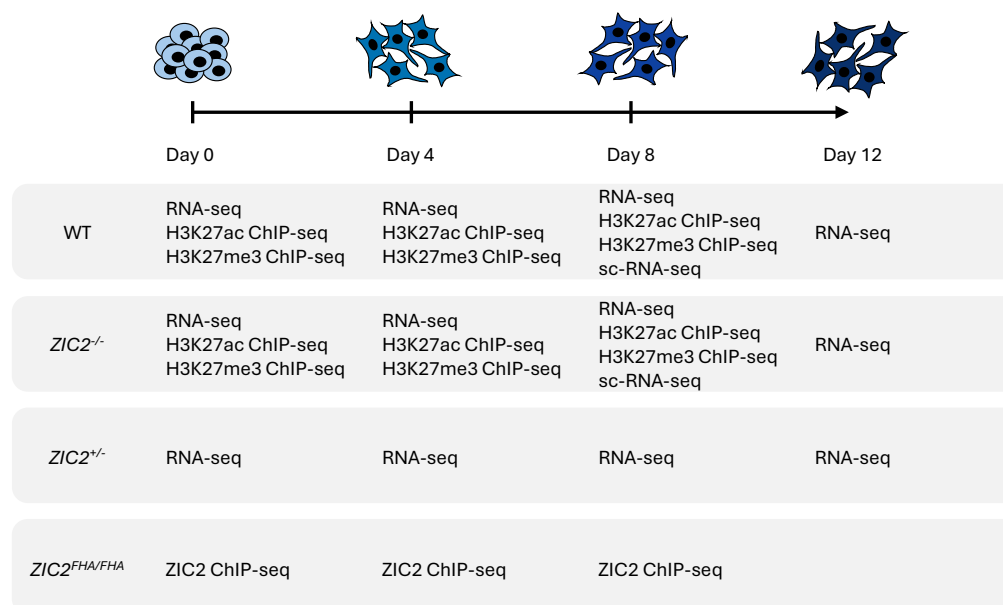
break at the C-terminal end of the *ZIC2* gene, we were able to insert a repair template containing a sequence coding for the Flag-HA tag. We confirmed the proper integration of the repair template by PCR followed by Sanger sequencing (Figure 4.4C). Two clones (#5 and #17) carrying the Flag-HA insertion in a homozygous manner were isolated. By western blot, we confirmed that ZIC2 could be specifically recognized with an anti-HA antibody (Figure 4.4D). In addition, ChIP-qPCR experiments using an anti-HA antibody were performed in WT and *ZIC2<sup>FHA/FHA</sup>* lines. Specific enrichment of ZIC2 binding in *ZIC2<sup>FHA/FHA</sup>* at a ZIC2 binding site located within a *LMX1B* intron (this binding site was predicted based on ChIP-seq data of mouse *Zic2*, generated in mESC by another member of our lab (Dr. María Mariner Faulí)) was confirmed (Figure 4.4E). Next, we verified that the generated *ZIC2<sup>FHA/FHA</sup>* hiPSC could differentiate properly into anterior neural progenitor cells by measuring the expression of key differentiation markers (i.e. *ZIC2*, *PAX6* and *LMX1A*) (Figure 4.4F). We also compared the ChIP-seq binding profiles of ZIC2 in hiPSC generated with either the anti-HA or the anti-ZIC2 antibody. Although the binding profiles were similar, peaks that were absent in both the *ZIC2<sup>FHA/FHA</sup>* (using the anti-HA antibody) and the *ZIC2<sup>-/-</sup>* (using the anti-ZIC2 antibody) cells could still be detected in WT cells using the commercial anti-ZIC2 antibody (Figure 4.4G). Therefore, the generated *ZIC2<sup>FHA/FHA</sup>* hiPSC lines enabled us to accurately map the binding profiles of ZIC2 in hiPSC and neural progenitors, eliminating potential cross-reactivity observed when using the anti-ZIC2 antibody.



**Figure 4.4: Generation of  $ZIC2^{FHA/FHA}$  hiPSC lines as a tool to generate specific ZIC2 binding profiles.**

(A) Examples of ChIP-seq tracks at the *WNT8B* and *WNT10B/WNT1* loci. ChIP experiments were performed using an anti-ZIC2 antibody in WT and  $ZIC2^{-/-}$  hiPSC. The red squares indicate ChIP-seq peaks detected in  $ZIC2^{-/-}$  cells. (B) Schematic representation of the CRISPR-Cas9 strategy used to add a Flag-HA tag at the C-terminus of ZIC2. The repair template is composed of a left and a right homology arm recognizing the C-terminal end of ZIC2 (excluding the endogenous STOP codon) and the first 50bp of the 3'UTR, respectively. In between there are sequences coding for a linker (Glycine-Serine-Glycine) followed by the Flag-HA epitopes and a STOP codon. The gRNA (purple scissors) is targeting ZIC2 in the C-terminus. Primers 1 and 2 used to genotype the isolated clones generated are also shown as orange arrows. (C-G) Validation of the  $ZIC2^{FHA/FHA}$  hiPSC lines obtained (namely, FHA5 and FHA17). (C) The proper integration of the repair template was confirmed by PCR followed by Sanger sequencing. The red squares indicate the bands that were cut from the gel and purified before Sanger sequencing. (D) The specific tagging of the ZIC2 protein was verified by western blot using both anti-HA and anti-ZIC2 antibodies in WT and  $ZIC2^{FHA/FHA}$  hiPSC lines. (E) ChIP-qPCR experiments were performed in WT and  $ZIC2^{FHA/FHA}$  lines using an anti-HA antibody to confirm the specific immunoprecipitation of ZIC2. ChIP enrichments were measured at negative control regions (chr4 and chr8) and within the *LMX1B* intron, where ZIC2 binding was predicted based on *Zic2* ChIP-seq data previously generated in mESC by our group. Enrichments were calculated as percentage of input. (F) The FHA17 line was differentiated into AntNPC, and the expression of key differentiation markers was measured by RT-qPCR on day 8. Expression levels were calculated using the  $2^{-\Delta C_t}$  method and the standard variation of technical triplicate is represented as error bars. (G) ChIP-seq tracks of ZIC2 at *WNT8B* and *WNT10B/WNT1* loci using anti-HA and anti-ZIC2 antibodies in WT and  $ZIC2^{-/-}$  hiPSC. The red square indicates a ZIC2 peak detected using the commercial anti-ZIC2 antibody but not with the anti-HA antibody. Tracks figures were generated using Figeno (Sollier et al., 2024).

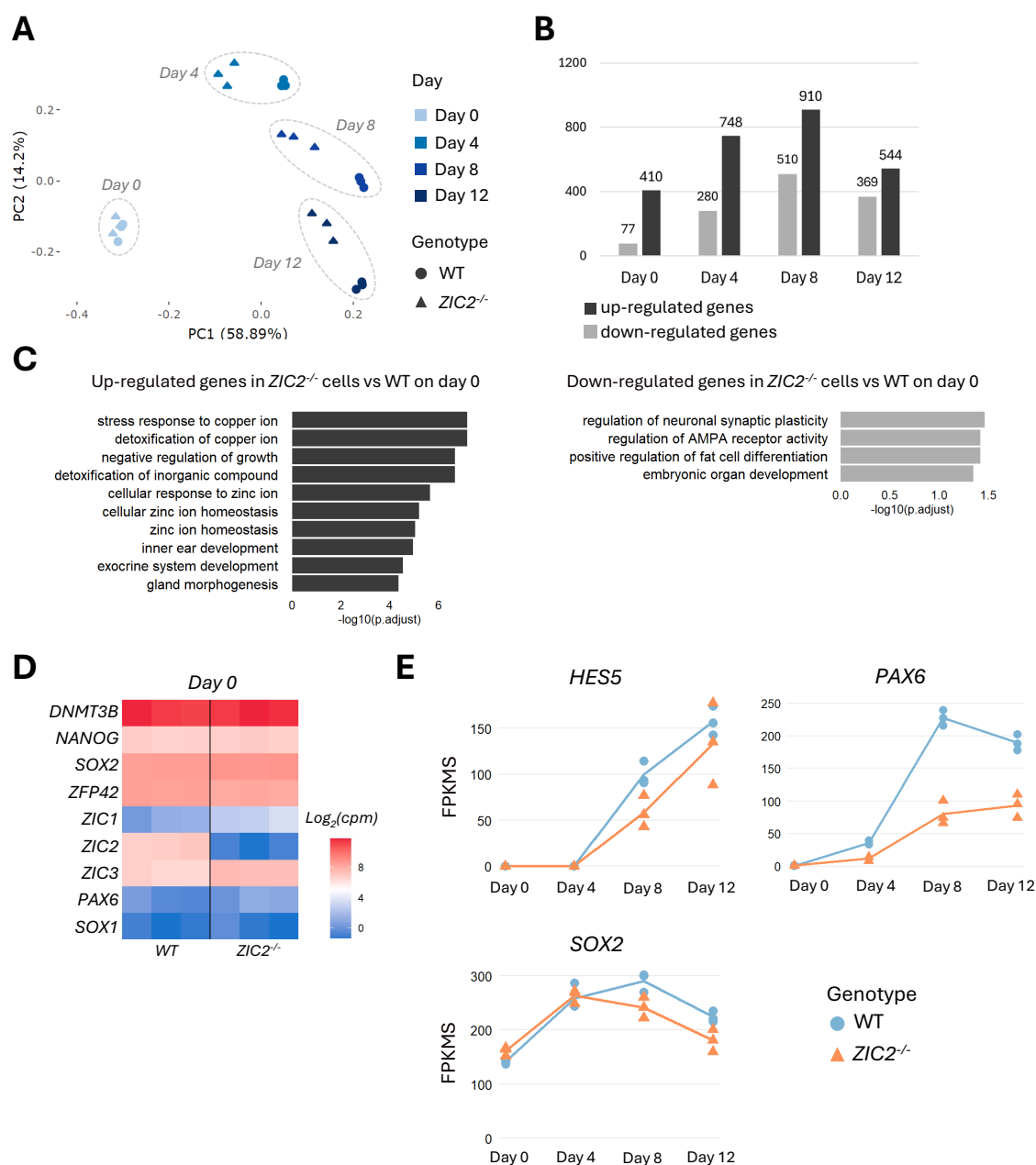
Thanks to cell lines generated and the differentiation protocol of hiPSC into AntNPC, we were able to generate several genomic datasets (RNA-seq, ChIP-seq and scRNA-seq) during the establishment of anterior neural identities, in WT,  $ZIC2^{-/-}$  and  $ZIC2^{+/-}$  cells (Figure 4.5). For the bulk RNA-seq experiments, samples were collected in triplicate, using three different hiPSC clonal lines for both  $ZIC2^{-/-}$  and  $ZIC2^{+/-}$  conditions. ChIP-seq experiments for chromatin marks (H3K27ac and H3K27me3) were performed in duplicate using  $ZIC2^{-/-}$  #14 and WT hiPSC lines. The ChIP-seq experiments for ZIC2 were also performed in duplicate using the  $ZIC2^{FHA/FHA}$  #17 hiPSC clonal line. For the single-cell RNA-seq experiments, the samples were derived from  $ZIC2^{-/-}$  #14 and WT cells.



**Figure 4.5: Overview of the genomic datasets generated in WT, *ZIC2*<sup>FHA/FHA</sup>, *ZIC2*<sup>-/-</sup> and *ZIC2*<sup>+/-</sup> cell lines.**

### 4.3. Characterization of molecular defects in *ZIC2*<sup>-/-</sup> cells

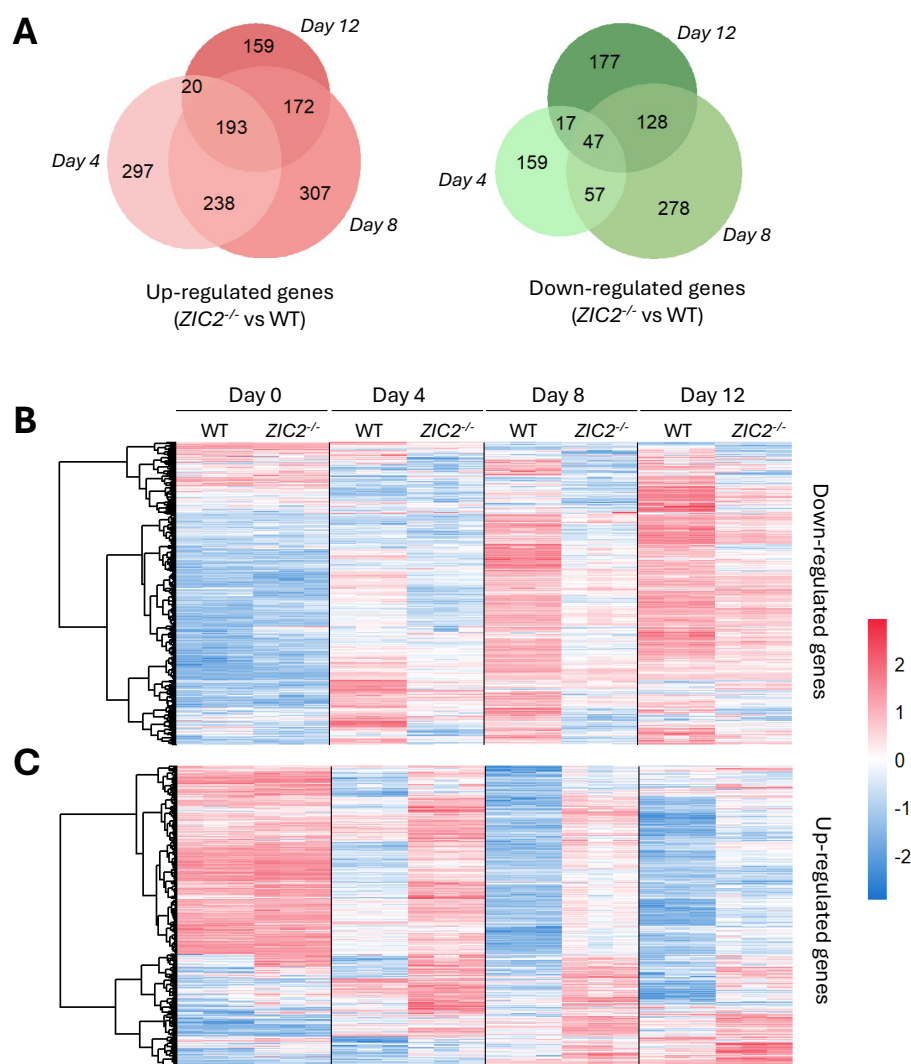
To start characterizing the molecular defects associated with *ZIC2* loss of function, we made use of the RNA-seq dataset generated in WT and *ZIC2*<sup>-/-</sup> cells throughout the AntNPC differentiation. A principal component analysis (PCA) showed that PC1 captured 58.9% of the variability and separated the samples according to their differentiation status, with day 0 samples located on the left of the plot and differentiated samples (day 4 to day 12) being on the right. Interestingly, on day 0, WT and *ZIC2*<sup>-/-</sup> samples clustered together while from day 4, WT and *ZIC2*<sup>-/-</sup> separated further from each other (Figure 4.6A). This is also reflected in the number of differentially expressed genes (DEGs) in *ZIC2*<sup>-/-</sup> vs WT, that is the lowest on day 0 and highest on day 8 (Figure 4.6B). GO term analysis of the DEGs on day 0 was performed and relatively poor enrichments were observed for terms that, for the most part, were unrelated to either pluripotency or neural differentiation (Figure 4.6C). Gene expression of pluripotency markers (*DNMT3B*, *NANOG*, *SOX2*, *ZFP42*) were checked in both WT and *ZIC2*<sup>-/-</sup> hiPSC, and no difference was observed between the two genotypes (Figure 4.6D). Moreover, no morphological or proliferation differences were noticed between the two hiPSC lines. Next, we decided to check whether the loss of *ZIC2* would impair the capacity of hiPSC to differentiate into neural progenitors. We therefore looked at the expression of neural progenitor markers (*SOX2*, *HES5* and *PAX6*) (Tchieu et al., 2017). Although the expression level of *PAX6* in *ZIC2*<sup>-/-</sup> cells were moderately reduced (2-fold), all of them were strongly induced in both WT and *ZIC2*<sup>-/-</sup> hiPSC upon differentiation into AntNPC (Figure 4.6E). The reduced expression level of *PAX6* could indicate a possible patterning defect (Stoykova et al., 2000), as discussed in following sections. We therefore concluded that the loss of *ZIC2* in hiPSC does not have major impacts on pluripotency and their capacity to differentiate into neural progenitors.



**Figure 4.6: The capacity of *ZIC2*<sup>-/-</sup> cells to differentiate into neural progenitor cells does not seem to be impaired.**

(A) Principal component analysis of the RNA-seq samples generated in WT and *ZIC2*<sup>-/-</sup> at the different time points of the AntNPC differentiation (day 0, day 4, day 8 and day 12). (B) Number of differentially expressed genes (DEGs) in *ZIC2*<sup>-/-</sup> vs WT cells across timepoints. To be considered as differentially expressed, a gene has to meet the following criteria: FPKM > 1 in at least one condition,  $|\log_2(\text{FoldChange})| > 1$  and  $p\text{-adj} < 0.01$ . (C) GO term analysis of the genes up and down-regulated in *ZIC2*<sup>-/-</sup> hiPSC were performed using *enrichGO* from *clusterProfiler* R package (Wu et al., 2021). Enrichments are shown as  $-\log_{10}(p\text{-value adjusted})$ . (D) Heatmap of gene expression dynamics (RNA-seq data) of different pluripotency genes, neural markers and *ZIC* family genes, in WT and *ZIC2*<sup>-/-</sup> hiPSC. The color scale is shown as  $\log_2(\text{counts per million})$ . (E) Expression in FPKMs of key neural markers (*HES5*, *PAX6* and *SOX2*) in WT (orange) and *ZIC2*<sup>-/-</sup> (light blue) cells during differentiation.

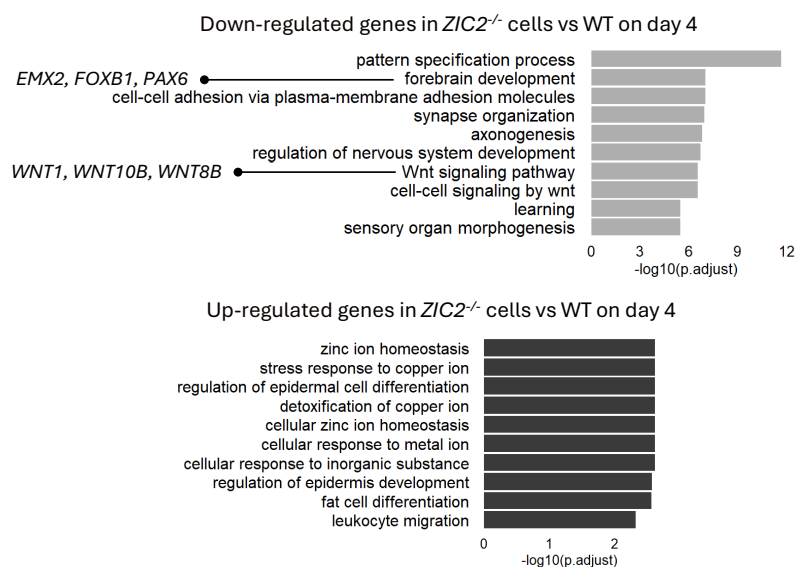
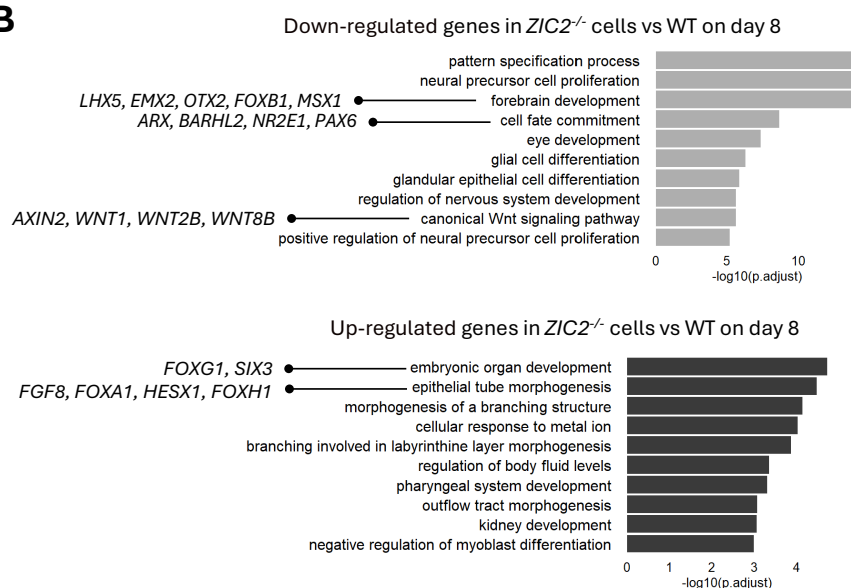
To further analyze the defects observed in *ZIC2*<sup>-/-</sup> neural progenitors, a differential expression analysis of *ZIC2*<sup>-/-</sup> vs WT cells was performed using DESeq from DESeq2 R package (Love et al., 2014). The overlap between the differential expressed genes (DEGs) at the investigated differentiation time points (day 4, day 8 and day 12) was calculated (Figure 4.7A). There is a large overlap of the DEGs across the different time points, with a higher number of DEGs found on day 8, for both up and down-regulated genes. Of all the genes considered as differentially expressed across the whole AntNPC differentiation, 80% can be identified on day 8, while the additional DEGs found only on day 12 might be the result of more indirect/secondary effects associated with the loss of *ZIC2*. In addition, the heatmaps of the expression of the DEGs (Figure 4.7B and C), revealed that changes initiated on day 4 are fully manifested already by day 8. For this reason, we decided to focus our next analysis on day 4 and day 8.



**Figure 4.7: Heatmaps of the differentially expressed genes.**

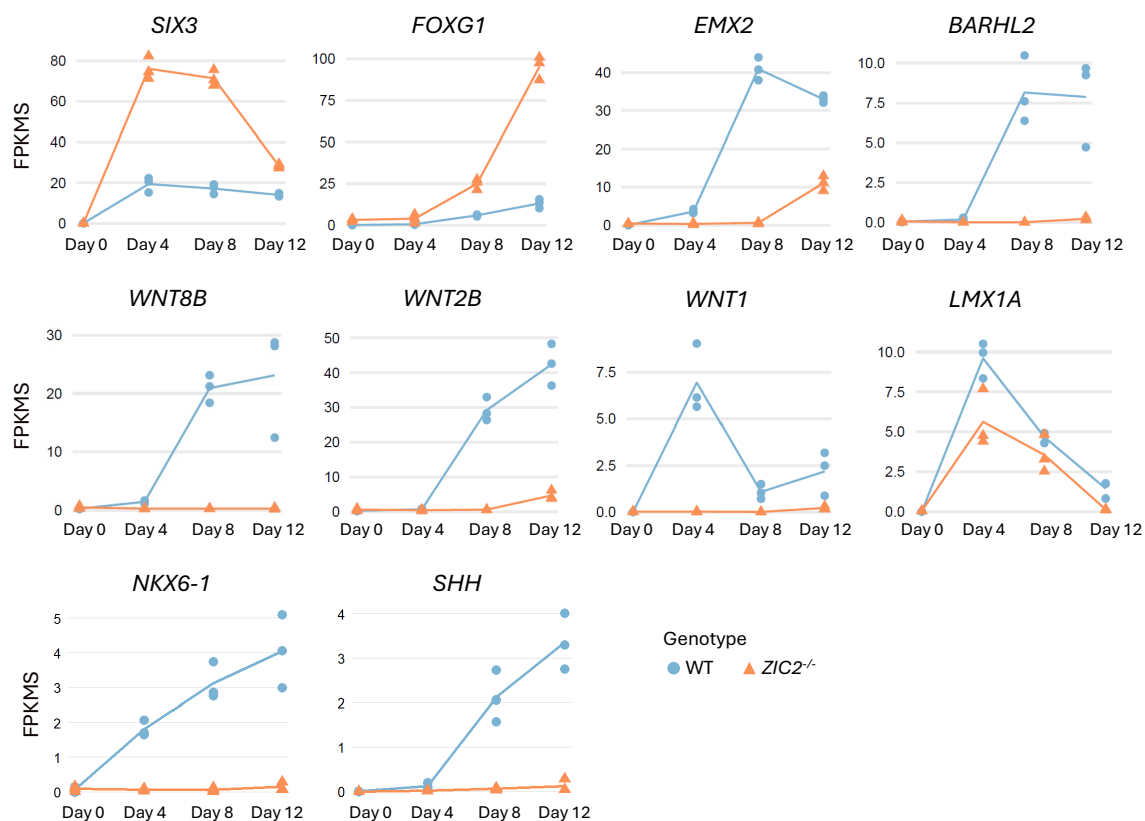
(A) Venn diagrams representing the overlap of the DEGs across timepoints (day 4 to day 12). (B-C) Heatmaps representing the expression of the down-regulated (B) and up-regulated (C) genes (day 4, day 8 and day 12 combined) in WT and *ZIC2*<sup>-/-</sup> cells on day 0, day 4, day 8 and day 12. Expression is shown as log(cpm) and scaled across genes (mean of 0 and variance of 1). Genes were clustered using the ward.D2 method.

Next, we conducted a GO term analysis of the DEGs on either day 4 or day 8 (Figure 4.8B and 4.8C, respectively). No particular terms were enriched for the up-regulated genes on day 4 while on day 8 the terms retrieved were somewhat general, referring to development and morphogenesis. Nevertheless, the genes upregulated on day 8 included major regulators of the rostral telencephalon, such as *SIX3*, *FOXP1* and *FGF8*. On the other hand, the GO term analysis of the down-regulated genes (both on day 4 and day 8) revealed more informative terms related to forebrain development, patterning and WNT signaling. When checking the genes associated with those terms, two key findings stood out. First, markers of the diencephalon and posterior telencephalon (*EMX2*, *OTX2*, *FOXB1* or *BARHL2*) (Bulfone et al., 2000; Parish et al., 2016; Suda et al., 2001) were down-regulated while, as mentioned above, markers of the rostral telencephalon (*FOXP1*, *SIX3* or *FGF8*) (Hébert & Fishell, 2008b; Sato et al., 2017) were up-regulated. Second, WNT signaling (*WNT1*, *WNT2B* or *WNT8B*) (Chizhikov & Iskusnykh, 2025; Iskusnykh et al., 2023) was downregulated. We therefore evaluated the expression dynamics of additional genes known to be involved in the previous developmental processes and signaling pathways and confirmed that markers of the posterior telencephalon/diencephalon were down-regulated (*EMX2*, *BARHL2*) in favor of more anterior identities (*SIX3*, *FOXP1*) (Figure 4.9). Moreover, the markers of dorsal identities resembling the cortical hem were either lost (*WNT2B*, *WNT8B*, *WNT1*) or down-regulated (*LMX1A*) in *ZIC2*<sup>-/-</sup> cells (Figure 4.9). Thus, these bulk gene expression changes suggest that upon loss of *ZIC2*, the positional identities of the AntNPC might be shifted towards the most anterior forebrain (i.e. rostral telencephalon), while more posterior (i.e. caudal telencephalon and diencephalon) and dorsal (i.e. cortical hem) identities might be lost.

**A****B**

**Figure 4.8: *ZIC2*<sup>-/-</sup> neural progenitors showed increased expression of rostral telencephalon markers and reduced levels of WNT signaling genes.**

(A–B) GO term analysis of the genes up and down-regulated in *ZIC2*<sup>-/-</sup> neural progenitors on day 4 (A) and day 8 (B) were performed using enrichGO from clusterProfiler R package. Representative genes associated with selected GO terms are highlighted. Enrichments are shown as  $-\log_{10}(\text{p-value adjusted})$ .

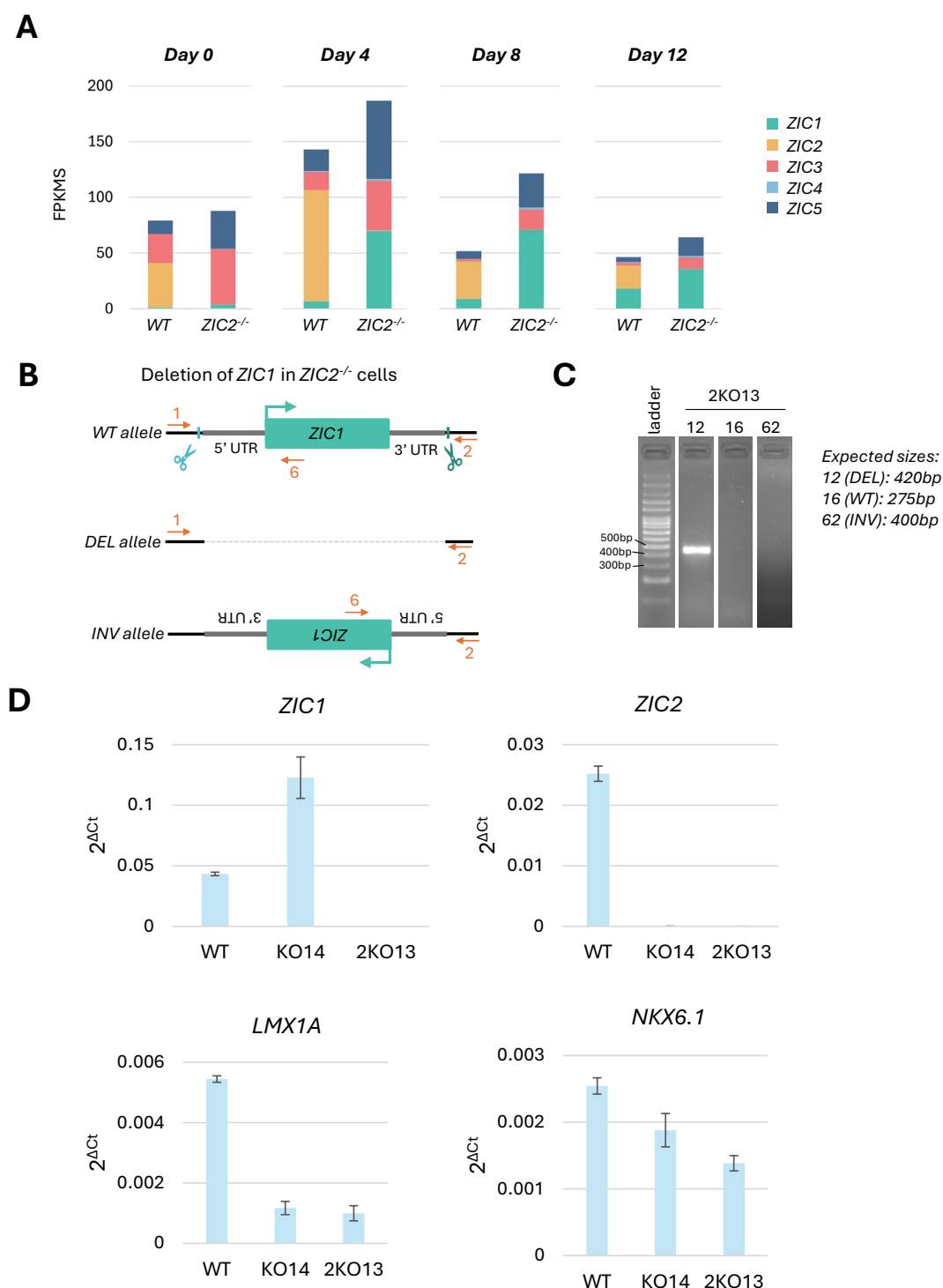


**Figure 4.9: Expression dynamics of key target genes in WT and *ZIC2*<sup>-/-</sup> cells.**

Gene expression dynamics of representative markers of the rostral telencephalon (*SIX3*, *FOXG1*), posterior telencephalon (*EMX2*), diencephalon (*EMX2*, *BARHL2*), cortical hem (*WNT8B*, *WNT2B*, *WNT1*, *LMX1A*), and ventral identity markers (*NKX6-1*, *SHH*) in WT (orange) and *ZIC2*<sup>-/-</sup> (light blue) cells.

#### 4.4. ZIC genes are strongly up-regulated in *ZIC2*<sup>-/-</sup> neural progenitors compared to WT

Surprisingly, ventral identity markers such as *NKX6.1* and *SHH* (Briscoe et al., 2000; Moreno-Bravo et al., 2010), were slightly induced in WT AntNPC but not in *ZIC2*<sup>-/-</sup> cells (Figure 4.9). This observation goes in opposition with the expected role of *ZIC2* as a dorsalization factor and the strong downregulation of cortical hem markers in *ZIC2*<sup>-/-</sup> cells. When examining the expression of all *ZIC* family genes together, we noticed that the overall *ZIC* expression levels were more than doubled in *ZIC2*<sup>-/-</sup> (121 FPKMs) neural progenitors on day 8, compared to WT (52 FPKMs). Notably, the upregulation of *ZIC1* (9 FPKMs in WT and 71 FPKMs in *ZIC2*<sup>-/-</sup> on day 8) largely contributed to the overall increase of *ZIC* expression in *ZIC2*<sup>-/-</sup> cells (Figure 4.10A). We therefore hypothesize that *ZIC* up-regulation in *ZIC2*<sup>-/-</sup> cells could be responsible, at least partially, for the defects observed upon *ZIC2* loss and, particularly, of the downregulation of neural ventral markers (e.g. *NKX6.1* and *SHH*). This hypothesis was further supported by the fact that *ZIC* proteins display high homology and partial functional redundancy (Houtmeyers et al., 2013; Inoue et al., 2007). Moreover, key developmental genes, such as *ZIC2*, exhibit high dosage sensitivity: not only can too little protein be deleterious for the cells, but also too high concentration can have detrimental effects. To test whether the defects observed in *ZIC2* knock-out cells could be due to an “over-compensation” mechanism by the other members of the *ZIC* family, we decided to knock-out *ZIC1* in a *ZIC2*<sup>-/-</sup> background using CRISPR-Cas9 (Figure 4.10B). A *ZIC1/ZIC2* double knock-out hiPSC line was isolated (Figure 4.10C) and differentiated into AntNPC in order to evaluate whether the loss of *ZIC1* could rescue, at least partially, the gene expression defects associated with *ZIC2* loss. Expression of key genes was measured by RT-qPCR and no major differences were observed between the single and the double knock-out cell lines (Figure 4.10D). We therefore concluded that the “over-compensation” hypothesis was not valid, and that the loss of the ventral identity markers in *ZIC2*<sup>-/-</sup> cells might be explained by the shift towards the most anterior identities (i.e. rostral telencephalon), which do not express high levels of neural ventral genes (Lagutin et al., 2003), rather than a loss of ventral identities.



**Figure 4.10: ZIC genes are up-regulated in *ZIC2*<sup>-/-</sup> neural progenitors.**

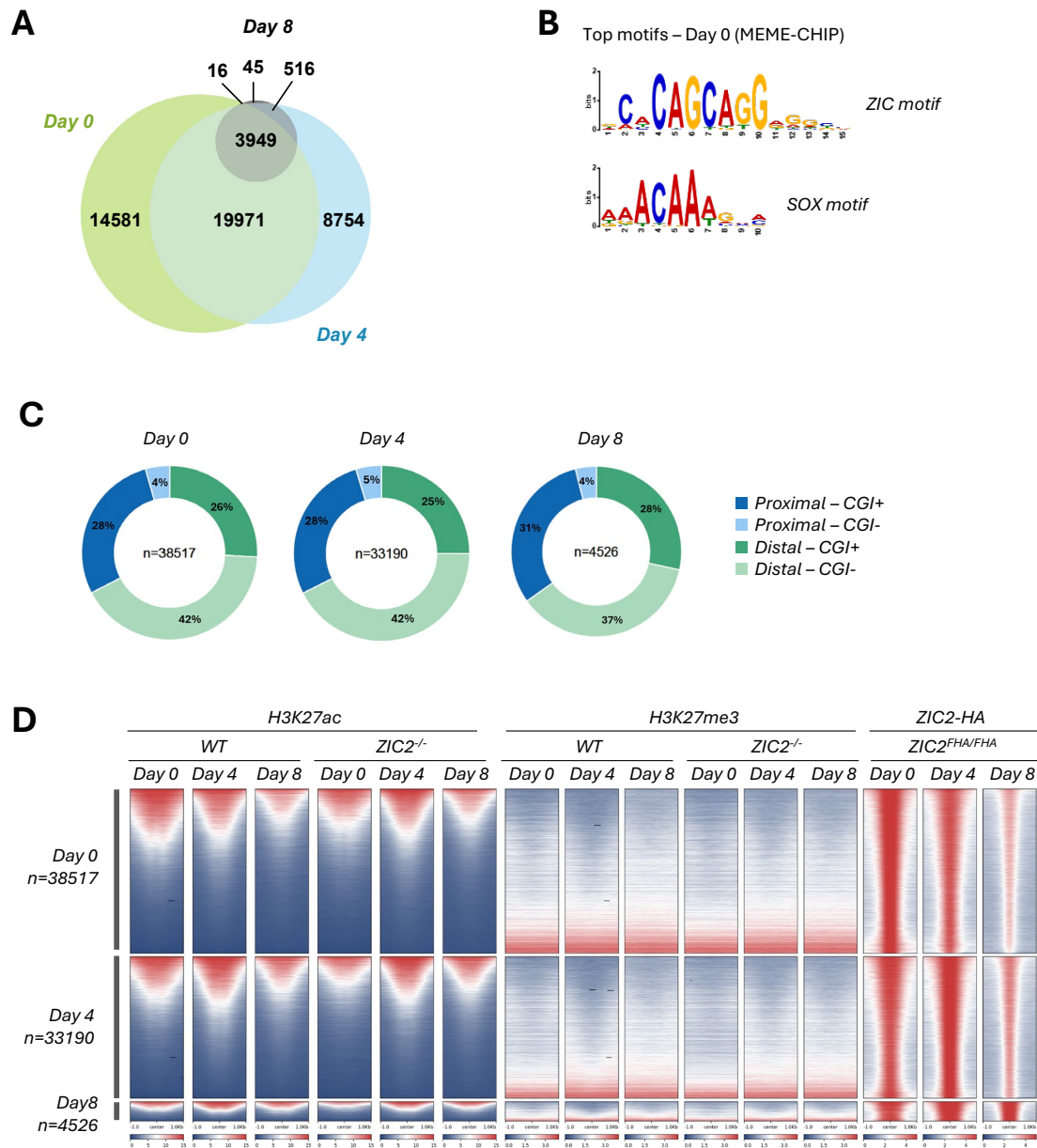
(A) Bar plots of ZIC family gene expression in FPKMs in WT and *ZIC2*<sup>-/-</sup> cells, at the different timepoints of the differentiation. (B) Schematic representation of the CRISPR-Cas9 strategy used to delete *ZIC1* gene in *ZIC2*<sup>-/-</sup> background generating a *ZIC1*<sup>-/-</sup>; *ZIC2*<sup>-/-</sup> double knock-out cell line. Primers (orange arrows) used to genotype the isolated clones generated are shown. (C) Genotyping PCR results of the isolated clone (2KO13). (D) Expression of key marker genes have been measured by RT-qPCR in WT, *ZIC2*<sup>-/-</sup> (KO14), *ZIC1*<sup>-/-</sup>; *ZIC2*<sup>-/-</sup> (2KO13) on day 8. Expression

levels were calculated using the  $2^{\Delta Ct}$  method and the standard variation of technical triplicate is represented as error bars.

#### 4.5. Characterization of ZIC2 binding profiles during AntNPC differentiation

To gain insights into the regulatory networks controlled by ZIC2 during neural induction and forebrain patterning, we generated the binding profiles of ZIC2, together with H3K27ac and H3K27me3 profiles, at several differentiation timepoints (day 0, day 4 and day 8). Already on day 0, ZIC2 binds extensively to the genome (38517 peaks). Of the 33190 peaks detected on day 4, 72% were already present on day 0 and 28% corresponded to new binding sites. On day 8, the number of ZIC2 peaks is strongly reduced (4526 peaks), of which less than 1% corresponds to new binding events (Figure 4.11A). Although ZIC2 is still notably expressed on day 8 (34 FPKMs), its expression levels are higher on both day 0 (40 FPKMs) and day 4 (99 FPKMs). Therefore, one possibility is that, as ZIC2 levels decrease on day 8, the total number of binding sites decreases, suggesting that its main regulatory functions might be executed on day 0 and/or day 4. Alternatively, we cannot exclude technical problems during the generation of the day 8 ZIC2 ChIP-seq data. To address this second possibility, we are planning to perform more replicates of the ZIC2 ChIP-seq experiments on day 8.

Motif discovery analysis of ZIC2 binding sites at the different timepoints was performed with MEME-CHIP (Machanick & Bailey, 2011). At all timepoints, the two top motifs retrieved were the ZIC motif, confirming the quality of the ChIP-seq data, and the SOX motif, suggesting that ZIC2 might cooperate with TFs belonging to the SOX family during neural differentiation (Figure 4.11B). Next, we looked at the genomic distribution of the ZIC2 binding regions, considering their proximity (a threshold of 1kb has been used) to promoters (proximal or distal) and to CpG islands (CGI) (i.e., CGI+ or CGI-). At all timepoints, the distribution was similar: about one third of ZIC2 binding is located in proximity to promoters while two thirds of ZIC2 binding is distal. ZIC2 binding at promoters mostly occurs in CG-rich context, which is an expected feature of most promoters, while at distal binding sites, ZIC2 is able to bind both CG-rich and CG-poor regions (Figure 4.11C). Evaluation of H3K27ac and H3K27me3 ChIP-seq signals in WT and *ZIC2*<sup>-/-</sup> cells around the ZIC2 bound regions showed that ZIC2 is able to bind putative *cis*-regulatory elements (CREs) displaying either an active (H3K27ac high; H3K27me3 low) or poised/inactive (H3K27ac high; H3K27me3 high) state (Figure 4.11D).

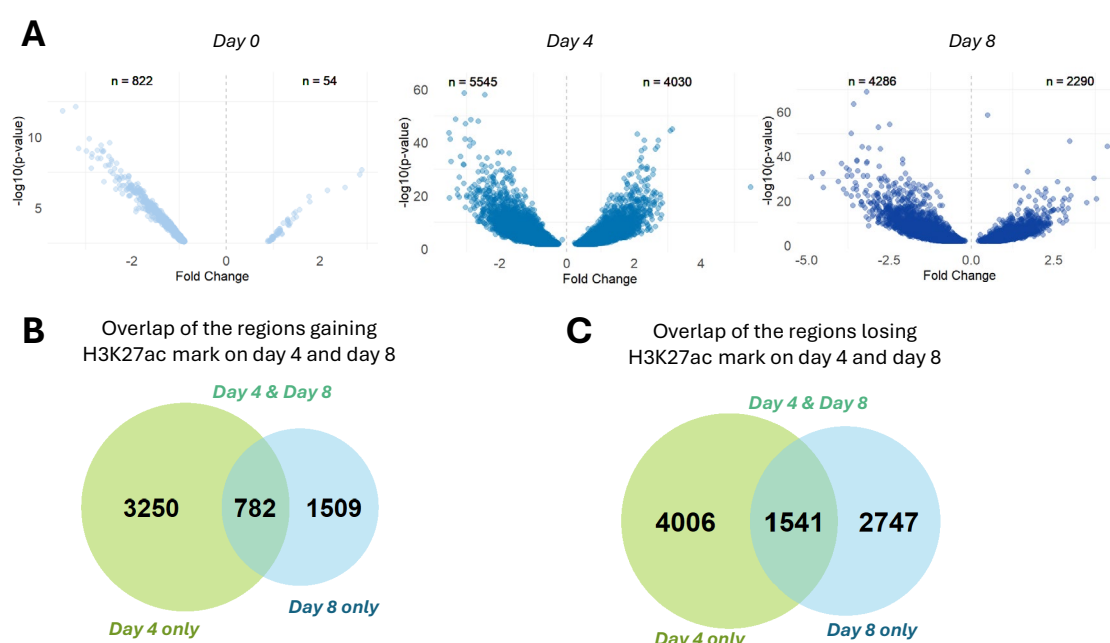


**Figure 4.11: ZIC2 binds both proximal and distal regions of the genome.**

(A) Venn diagram of ZIC2 binding dynamics during differentiation. The numbers indicated correspond to the number of ZIC2 peaks detected (macs2 peak calling). (B) Motif discovery analysis of ZIC2-bound regions. At any time point, ZIC and SOX binding motifs were retrieved. This analysis was performed using MEME-CHIP tool (Machanick & Bailey, 2011) and the top motifs identified when considering the day 0 ZIC2 peaks are shown. (C) Donut charts representing the distribution of ZIC2 binding regions depending on their distance to CGI and to promoter regions. A peak is considered proximal if its distance to the closest TSS is lower than 1kb. Similarly, a peak is considered CGI+, if its distance to the closest CGI region is lower than 1kb. (D) Heatmap of the ChIP-seq signals for ZIC2, H3K27ac and H3K27me3 at ZIC2-bound regions on day 0, day 4 and day 8.

Since H3K27ac is a typical mark of both active promoters and active enhancers, we then performed a differential binding analysis of H3K27ac in *ZIC2*<sup>-/-</sup> vs WT at the three different timepoints (Figure 4.12A) in order to identify CREs whose activity could be controlled by *ZIC2*. On day 0, relatively few regions showing significant differences in H3K27ac were found (822 and 54 sites lost or gained acetylation, respectively), in agreement with the minor transcriptional defects observed in *ZIC2*<sup>-/-</sup> hiPSC. In contrast, on day 4 and 8, H3K27ac differences between *ZIC2*<sup>-/-</sup> and WT cells were significantly more pronounced (Figure 4.12A), with more regions losing rather than gaining H3K27ac in *ZIC2*<sup>-/-</sup> cells.

Next, we compared the regions gaining or losing H3K27ac in *ZIC2*<sup>-/-</sup> cells on either day 4 or day 8. These analyses revealed that 14% of the regions gaining H3K27ac were shared between day 4 and day 8, while the remaining ones gained H3K27ac on day 4 only or on day 8 only (Figure 4.12B). Similarly, among the regions losing H3K27ac in *ZIC2*<sup>-/-</sup> cells, 15% were shared between day 4 and day 8 (Figure 4.12C). To better understand the functional relevance of these putative CREs and evaluate whether *ZIC2* directly controls their activity, we separated our analysis into two main parts: first, looking at the regions gaining H3K27ac (Figure 4.13 and 4.14) and, second, focusing on the regions losing H3K27ac (Figure 4.15 and 4.16).

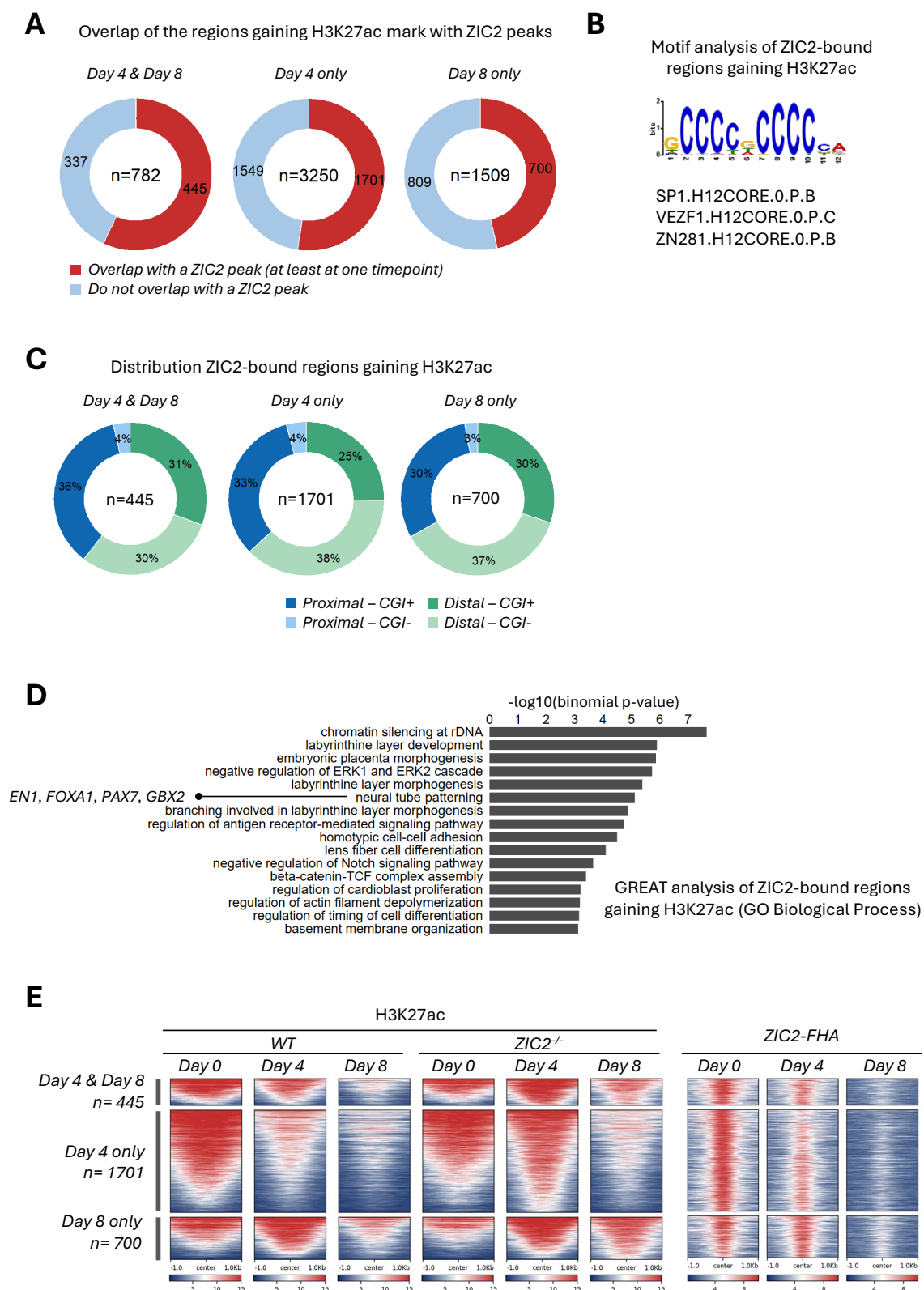


**Figure 4.12: Differential binding analysis of H3K27ac in *ZIC2*<sup>-/-</sup> vs WT cells throughout AntNPC differentiation.**

(A) Volcano plot representing differential binding ( $FDR \leq 0.05$ ) of H3K27ac in *ZIC2*<sup>-/-</sup> vs WT cells on day 0, day 4 and day 8. (B – C) Venn diagram representing the overlap of the regions gaining (B) or losing (C) H3K27ac mark on day 4 and day 8. These regions were divided into 3 subgroups: regions gaining/losing H3K27ac only on day 4 (Day 4 only), on day 8 (Day 8 only) or on both days (Day 4 & Day 8).

#### 4.6. Characterization of ZIC2 bound regions gaining H3K27ac in ZIC2<sup>-/-</sup> cells

Regarding the regions gaining H3K27ac, approximately half of them were bound by ZIC2 at least at one of the two timepoints (Figure 4.13A). Surprisingly, motif discovery analysis of the ZIC2-bound regions gaining H3K27ac failed to identify the ZIC motif (Figure 4.13B). Instead, a CG-rich motif was uncovered, which could be attributed to the promiscuous binding of ZIC2 to CGI, which typically displays high chromatin accessibility. In agreement with this, more than 60% of the ZIC2-bound regions gaining H3K27ac were proximal to CGI (Figure 4.13C). Moreover, GREAT analysis (McLean et al., 2010; Tanigawa et al., 2022) of these regions revealed enrichments in terms that were not particularly related to the observed transcriptional defects, with the exception of “neural tube patterning”. This term included regions potentially linked to genes like *EN1*, *FOXA1*, *PAX7* or *GBX2* (Figure 4.13D). Both *EN1* and *PAX7* are not expressed in our differentiation system in none of the conditions. *GBX2* and *FOXA1* are not expressed in WT cells at any timepoint but are up-regulated in ZIC2<sup>-/-</sup> cells from day 8, suggesting that ZIC2 might act as a repressor for these genes. Subsequently, the heatmap of the ZIC2-bound regions gaining H3K27ac in ZIC2<sup>-/-</sup> revealed that most of these regions are already active in WT and ZIC2<sup>-/-</sup> hiPSC (i.e. day 0) and then failed to be silenced in ZIC2<sup>-/-</sup> cells on day 4 and/or day 8 in comparison with WT cells. Accordingly, ZIC2 binding at regions gaining H3K27ac mark is generally stronger on day 0 than at later differentiation time points (Figure 4.13E). These observations suggest that, upon neural differentiation, ZIC2 might help repressing CREs that are active in hiPSC. However, since these regions were not enriched in the ZIC2 motif and were often found in a highly accessible CGI-rich context, they might not be directly regulated by ZIC2.

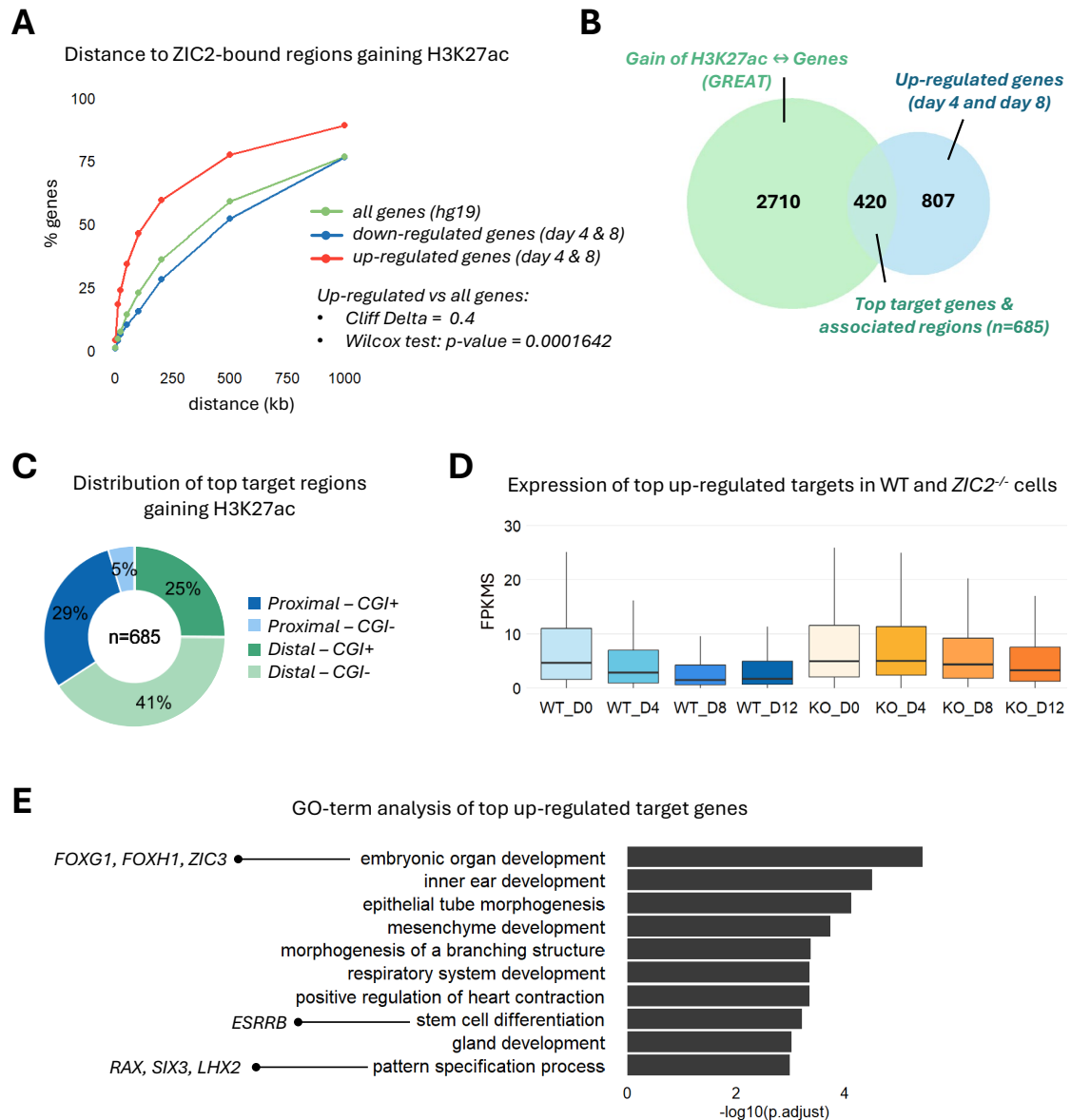


**Figure 4.13: Regions gaining H3K27ac in ZIC2<sup>-/-</sup> cells.**

(A) Donut charts representing the regions gaining H3K27ac that overlap (red) or not (blue) with a ZIC2 peak at least at one time point. (B) Motif discovery analysis of the ZIC2-bound regions gaining H3K27ac. This analysis was performed using MEME-CHIP tool. (C) Donut charts representing the distribution of the ZIC2-bound regions gaining H3K27ac based on their distance to promoters and

*CGI. A peak was considered proximal if its distance to the closest TSS was lower than 1kb. Similarly, a peak was considered CGI+, if its distance to the closest CGI region as lower than 1kb. (D) GO enrichment analysis of the ZIC2-bound regions gaining H3K27ac. Analysis was performed with GREAT (McLean et al., 2010) and enrichments are shown as  $-\log_{10}(\text{binomial } p\text{-value})$ . (E) Heatmap of H3K27ac ChIP signals at the ZIC2-bound regions gaining H3K27ac in ZIC2<sup>-/-</sup> cells. The heatmap was subdivided as defined in Figure 4.12B.*

To more directly interrogate whether the ZIC2-bound regions gaining H3K27ac in ZIC2<sup>-/-</sup> cells could represent CREs repressed by ZIC2, we evaluated their potential effects on gene expression on day 4 and day 8. For each gene, we calculated the distance to the closest ZIC2-bound region gaining H3K27ac in ZIC2<sup>-/-</sup> cells and represented those distances as cumulative distribution considering three groups of genes: genes that are up-regulated in ZIC2<sup>-/-</sup> cells, genes that are down-regulated in ZIC2<sup>-/-</sup> cells and all genes. Up-regulated genes were significantly closer to ZIC2-bound regions gaining H3K27ac compared to the other two groups, indicating that ZIC2 could repress those regions and, in turn, silence nearby genes (Figure 4.14A). We therefore defined a subset of top target up-regulated genes and their associated ZIC2-bound regions using the following strategy: ZIC2-bound regions gaining H3K27ac in ZIC2<sup>-/-</sup> cells were associated with their putative target genes using GREAT (basal extension plus mode) and then overlapped with those genes that were up-regulated in ZIC2<sup>-/-</sup> cells (day 4 and day 8 combined) (Figure 4.14B). As expected, the distribution of the top target regions gaining H3K27ac was similar to the overall distribution of the ZIC2-bound regions gaining H3K27ac (i.e. approximately 60% were close to CGI and >30% were close to promoter regions) (Figure 4.14C). We checked the expression levels of the top-up regulated genes and confirmed that they mainly correspond to genes that are expressed on day 0 and then progressively become silenced during the differentiation. However, in ZIC2<sup>-/-</sup> cells, the silencing of these genes is delayed (Figure 4.14D). GO analysis of the top target up-regulated genes revealed rather weak enrichments for terms related to non-neural developmental processes (Figure 4.14E). Nevertheless, these top target genes included major rostral telencephalon regulators, such as, *FOXP1*, *RAX* or *SIX3*, which were already described in previous sections.



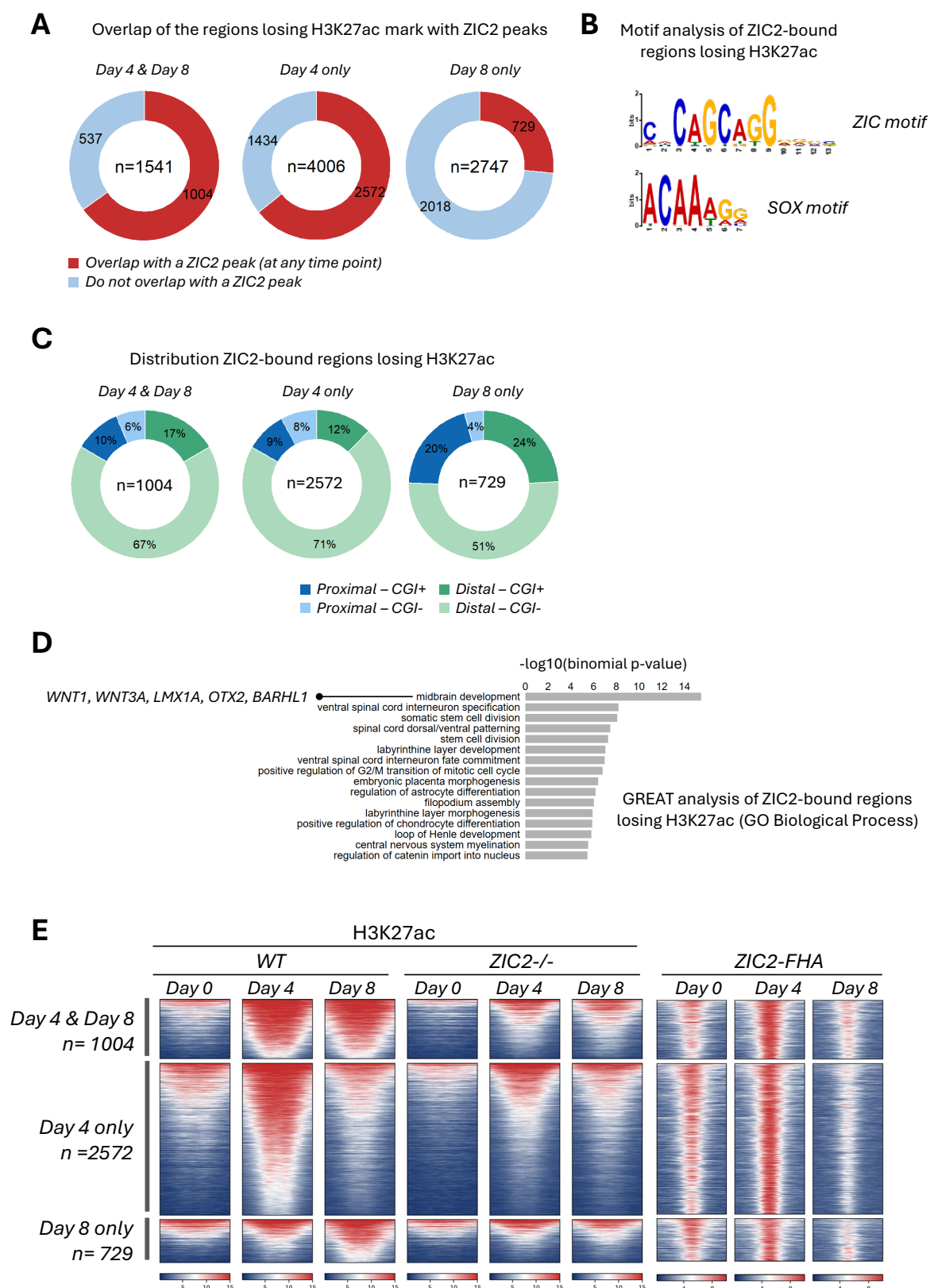
**Figure 4.14: ZIC2-bound regions gaining H3K27ac in ZIC2<sup>-/-</sup> cells are associated with genes whose silencing upon pluripotency exit is delayed ZIC2<sup>-/-</sup> cells.**

(A) The distance of each gene to the closest ZIC2-bound peak gaining H3K27ac in ZIC2<sup>-/-</sup> cells was calculated and those distances are represented as a cumulative distribution, considering three gene groups: genes that are up-regulated ZIC2<sup>-/-</sup> cells, genes that are down-regulated in ZIC2<sup>-/-</sup> cells and all genes. (B) Venn diagram representing the overlap between the up-regulated genes in ZIC2<sup>-/-</sup> (day 4 and day 8 combined) and the genes associated with ZIC2-bound regions gaining H3K27ac in ZIC2<sup>-/-</sup> cells. ZIC2-bound regions were linked to putative target genes using GREAT (McLean et al., 2010) with the basal plus extension mode. The overlap defines the top target up-regulated genes and their associated regions. (C) Donut chart representing the distribution of the top target regions based on their distance to promoters and CGI. (D) Boxplots of the expression dynamics in FPKMS of the top up-regulated genes (as defined in B) in WT and ZIC2<sup>-/-</sup> cells during AntNPC differentiation. (E) GO term analysis of biological processes enriched among the top target genes up-regulated in ZIC2<sup>-/-</sup> cells was performed using enrichGO from clusterProfiler R package. Enrichments are shown as  $-\log_{10}(\text{p-value adjusted})$ .

Taken altogether, these findings suggest that *ZIC2* might have a role as repressor of putative CREs that control the expression of pluripotency-associated genes. However, since many of those putative CREs were not enriched in the *ZIC2* motif and were often found in a highly accessible CGI-rich context, they might not be directly regulated by *ZIC2*.

#### 4.7. Characterization of *ZIC2* bound regions losing H3K27ac in *ZIC2*<sup>-/-</sup> cells

Next, we similarly analyzed the regions losing H3K27ac in *ZIC2*<sup>-/-</sup> cells. About 50% of all these regions overlapped with a *ZIC2* peak, although this percentage increased to 64% if we consider regions losing H3K27ac in day 4 only or in both day 4 and day 8 groups (Figure 4.15A). Importantly, MEME-ChIP analysis of these regions retrieved both the *ZIC* and *SOX* motifs (Figure 4.15B). In addition, the majority of these regions were distal from promoter regions (80% distal) and from CGI (>60% CGI-) (Figure 4.15C). Therefore, in contrast with the regions gaining H3K27ac, these findings support a direct role for *ZIC2* as an activator of distal CRE (i.e. enhancers). GREAT analysis of the *ZIC2*-bound regions losing H3K27ac revealed strong enrichments for several terms associated with brain development and patterning. Notably, these terms included genes like *BARHL1*, *LMX1A*, *WNT1* or *WNT3A*, all of them being down-regulated in *ZIC2*<sup>-/-</sup> neural progenitors, thus further supporting *ZIC2* as an activator of enhancers controlling the induction of important brain patterning regulators during neural differentiation (Figure 4.15D). In agreement with this possibility, examination of H3K27ac levels around the *ZIC2*-bound regions losing H3K27ac in *ZIC2*<sup>-/-</sup> cells revealed that, in WT cells, the majority of these regions were initially inactive on day 0 and then gained H3K27ac on day 4 and/or 8 (Figure 4.15E). Furthermore, although *ZIC2* was already bound to many of these regions on day 0, its binding levels increased considerably on day 4, before decreasing again on day 8. Notably, the majority of these regions (83%) failed to gain H3K27ac already in day 4 *ZIC2*<sup>-/-</sup> cells, while only a small fraction (17%) displayed H3K27ac defects on day 8 only (Figure 4.15E). Since the number of downregulated genes in *ZIC2*<sup>-/-</sup> cells is considerably higher on day 8 than on day 4, these observations suggest that defects in enhancer activation might precede transcriptional ones.

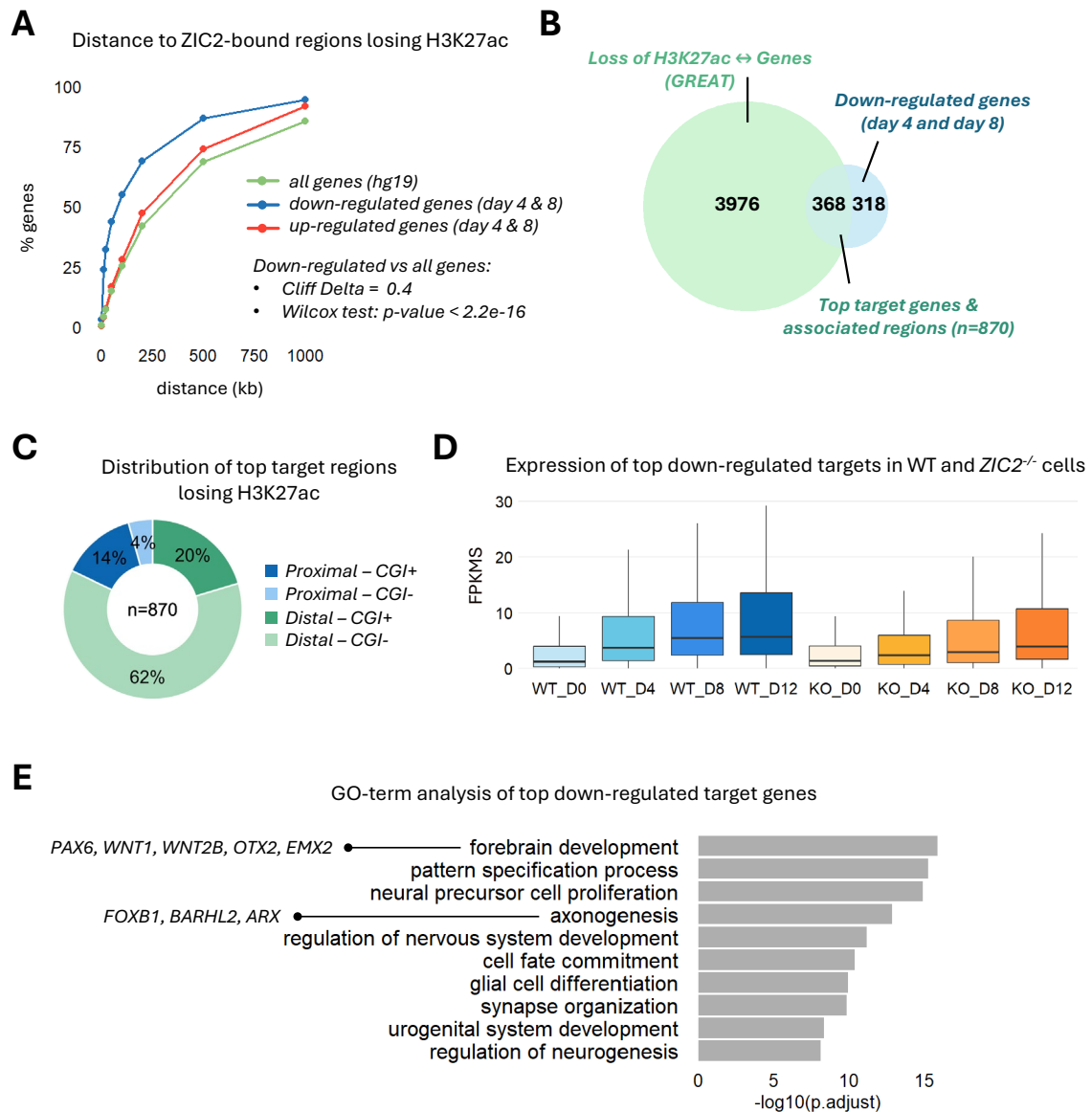


**Figure 4.15: Regions losing H3K27ac mark in ZIC2<sup>-/-</sup> cells.**

(A) Donut charts representing the regions losing H3K27ac in ZIC2<sup>-/-</sup> cells that overlap (red) or not (blue) with a ZIC2 peak. (B) Motif discovery analysis of the ZIC2-bound regions losing H3K27ac in ZIC2<sup>-/-</sup> cells. This analysis was performed using the MEME-CHIP tool. (C) Donut charts

*representing the distribution of the ZIC2-bound regions losing H3K27ac in ZIC2<sup>-/-</sup> cells, based on their distance to promoters and CGI. (D) GREAT analysis of the ZIC2-bound regions losing H3K27ac in ZIC2<sup>-/-</sup> cells. Enrichments are shown as  $-\log_{10}(\text{binomial } p\text{-value})$ . (E) Heatmap of the H3K27ac ChIP signals at the ZIC2-bound regions losing H3K27ac in ZIC2<sup>-/-</sup> cells. The heatmap was subdivided as defined in Figure 4.12C.*

Next, to further evaluate the potential role of ZIC2 as an activator of enhancers controlling the expression of genes involved in brain development and patterning, we evaluated the correlation between regions losing H3K27ac in ZIC2<sup>-/-</sup> cells and the expression levels of nearby genes. Indeed, as shown in the cumulative distribution plot, genes down-regulated in ZIC2<sup>-/-</sup> cells tend to be significantly closer to ZIC2-bound regions losing H3K27ac compared to all genes (Figure 4.16A). Using GREAT, we associated ZIC2-bound regions losing H3K27ac in ZIC2<sup>-/-</sup> cells with their putative target genes (basal plus extension mode) and calculated their overlap with the down-regulated genes in ZIC2<sup>-/-</sup> cells (day 4 and day 8 combined). The overlap defined the top down-regulated genes and their corresponding regions (top target regions losing H3K27ac) (Figure 4.16B). The majority of the top target regions (82%) were distal from gene promoters, thus in agreement with their role as enhancers (Figure 4.16C). Moreover, these regions were globally associated with genes whose expression was induced in WT cells but delayed or impaired in ZIC2<sup>-/-</sup> cells during neural differentiation (Figure 4.16D). The GO term analysis of the top down-regulated genes revealed strong enrichments for terms related to forebrain development, patterning and WNT signaling (Figure 4.16E). These results therefore depict ZIC2 as an important enhancer activator during forebrain development and patterning.

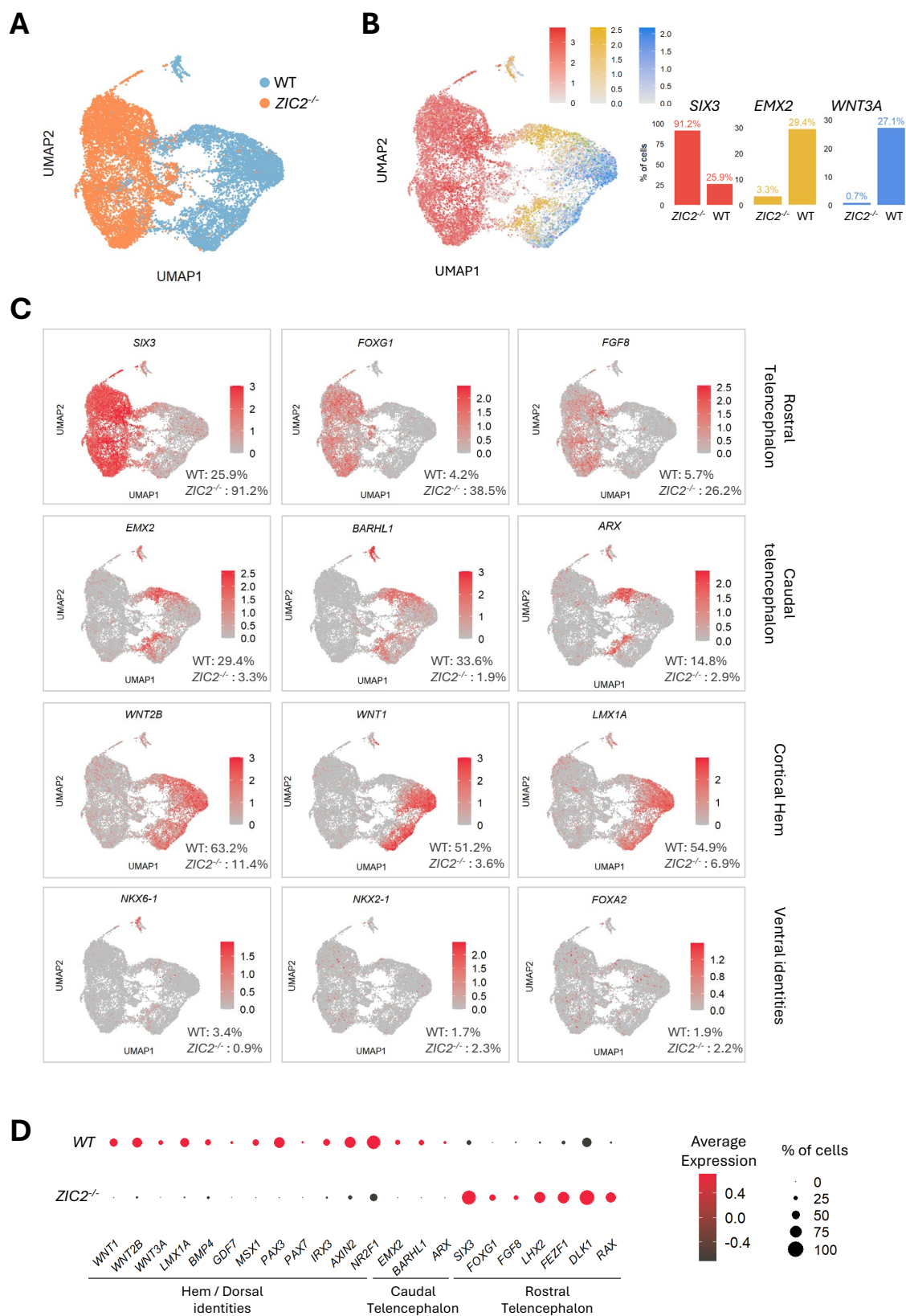


**Figure 4.16: ZIC2-bound regions losing H3K27ac in ZIC2<sup>-/-</sup> cells are associated with genes whose expression failed to be induced in ZIC2<sup>-/-</sup> AntNPC.**

(A) The distance of each gene to the closest ZIC2-bound peak losing H3K27ac in ZIC2<sup>-/-</sup> cells was calculated and is represented as a cumulative distribution, considering three gene groups: genes that are up-regulated ZIC2<sup>-/-</sup> cells, genes that are down-regulated in ZIC2<sup>-/-</sup> cells and all genes.. (B) Venn diagram representing the overlap between the down-regulated genes (day 4 and day 8 combined) and the genes associated with ZIC2-bound regions losing H3K27ac in ZIC2<sup>-/-</sup> cells. ZIC2-bound regions were linked to putative target genes using GREAT (McLean et al., 2010) with the basal plus extension mode. The overlap defines the top target down-regulated genes and their associated regions. (C) Donut chart representing the distribution of the top target regions losing H3K27ac in ZIC2<sup>-/-</sup> cells based on their distance to promoters and CGI. (D) Boxplot showing the expression levels (in FPKMS) of the top down-regulated genes in WT and ZIC2<sup>-/-</sup> cells during AntNPC differentiation. (E) GO term (Biological Processes) analysis of the top target down-regulated genes in ZIC2<sup>-/-</sup> cells was performed using enrichGO (clusterProfiler R package). Enrichments are shown as  $-\log_{10}(p\text{-value adjusted})$ .

#### 4.8. The positional identities of neural progenitors are shifted upon loss of *ZIC2*

To complete the picture and better understand the molecular defects associated with *ZIC2* loss, we conducted single-cell RNA-seq experiment including WT and *ZIC2*<sup>-/-</sup> AntNPC on day 8. This technique helped us to distinguish between the different cellular identities that are established during the AntNPC differentiation and to interpret better the transcriptional changes observed at the single-cell level. On the UMAP space, the WT and *ZIC2*<sup>-/-</sup> cells were segregated in two main clusters, confirming the pronounced transcriptional differences that emerge on day 8 AntNPC in the absence of *ZIC2* (Figure 4.17A). The expression of three markers of different forebrain regions (*SIX3* (rostral telencephalon), *EMX2* (posterior telencephalon/rostral diencephalon) and *WNT3A* (cortical hem)) that were mis-regulated target in *ZIC2*<sup>-/-</sup> cells according to the previous bulk RNA-seq analyses confirmed that, upon loss of *ZIC2*, the positional identities of neural progenitors are shifted towards the anterior-most forebrain identities, while dorsal ones (i.e. cortical hem) are lost (Figure 4.17B). To strengthen this conclusion, we plotted in the UMAP space several markers of the rostral telencephalon/anterior neural ridge (ANR) (*SIX3*, *FOXG1* and *FGF8*) (Hébert & Fishell, 2008b; Sato et al., 2017), caudal telencephalon/rostral diencephalon (*EMX2*, *BARHL1* and *ARX*) (Bulfone et al., 2000; Parish et al., 2016; Suda et al., 2001) and the cortical hem (*WNT2B*, *WNT1* and *LMX1A*) (Chizhikov & Iskusnykh, 2025; Iskusnykh et al., 2023). Importantly, all these genes displayed the same expression patterns previously described for *SIX3*, *EMX2* and *WNT3A*: i.e. gain of the most anterior identities and loss of the dorsal ones upon *ZIC2* loss (Figure 4.17C). These observations were further confirmed by evaluating the expression of forebrain patterning markers, as illustrated in the dot plot shown in Figure 4.17D.



**Figure 4.17: *ZIC2*<sup>-/-</sup> neural progenitor cells show a shift towards the most anterior neural progenitor identities.**

(A) UMAP plot of WT (light blue) and *ZIC2*<sup>-/-</sup> (orange) neural progenitors on day 8. (B) UMAP plot of WT and *ZIC2*<sup>-/-</sup> neural progenitors on day 8. Expression of *SIX3* (red), *EMX2* (yellow) and *WNT3A*

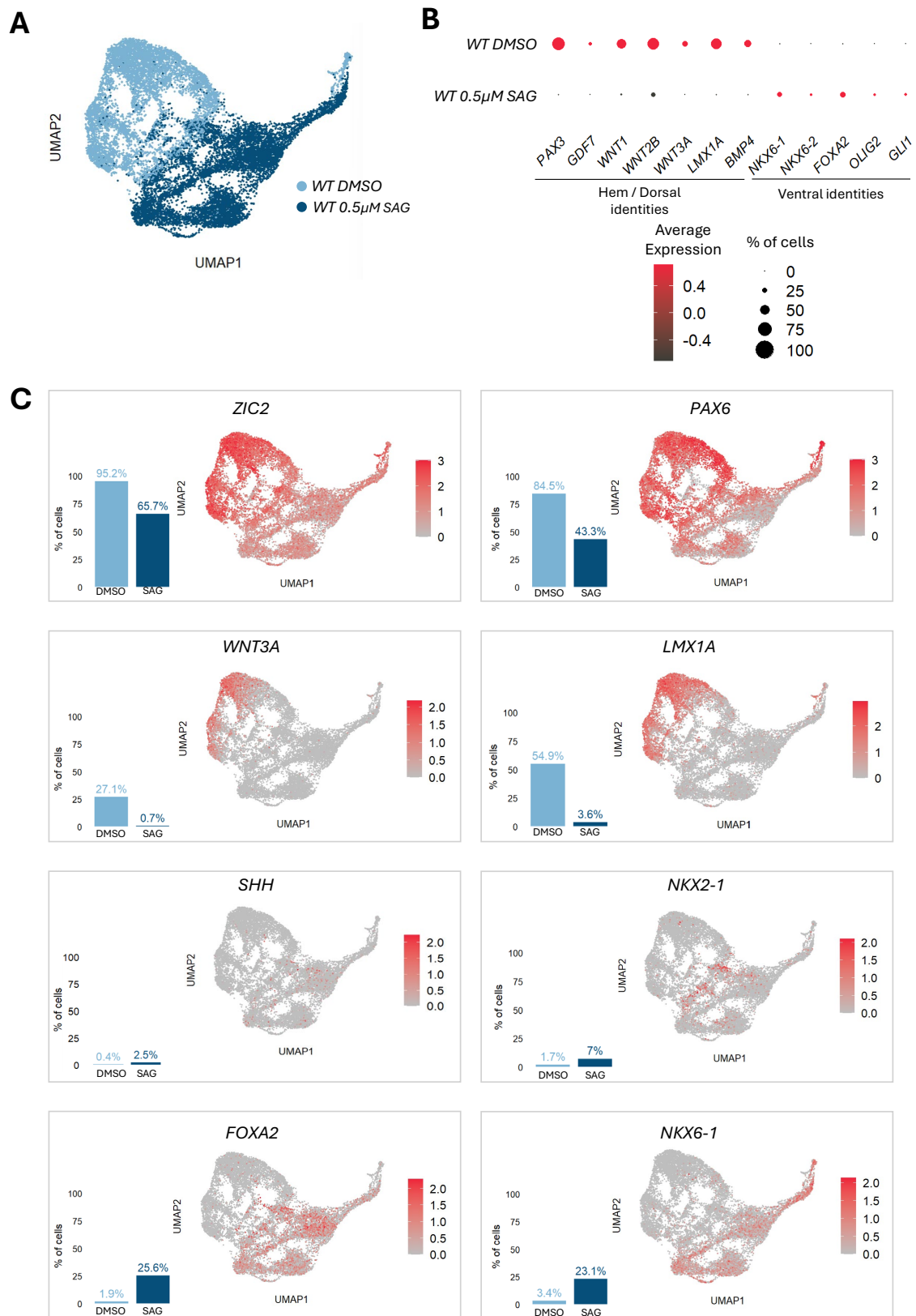
(blue) is highlighted. The percentage of cells expressing each marker in WT and *ZIC2*<sup>-/-</sup> is shown in the barplots. (C) UMAP plots of the expression of key target genes in WT and *ZIC2*<sup>-/-</sup> neural progenitors. Expression is calculated as  $\log_{10}(\text{feature counts per cell} / \text{total counts per cell}) \times 10000$ . (D) Dot plot representing the level of expression (color of the dot, z-score) and the percentage of cells expressing (size of the dot) the indicated forebrain patterning markers in WT and *ZIC2*<sup>-/-</sup> neural progenitors.

#### 4.9. Investigating how the loss of *ZIC2* affects the responsiveness to hedgehog signaling in neural progenitors

Although the expression of dorsal/cortical hem identity genes, like *LMX1A*, *WNT3A* or *WNT2B*, were consistently lost in *ZIC2*<sup>-/-</sup> neural progenitors, this was not accompanied by a significant increase in the number of cells expressing high levels of neural ventral markers (e.g., *NKX6-1*, *NKX2-1* and *FOXA2*) (Figure 4.17C). WNT and SHH signaling pathways antagonize each other and induce either dorsal or ventral identities during neural tube development, respectively (Rash & Grove, 2007; Ulloa & Martí, 2010). However, and in agreement with mouse *in vivo* expression profiles (Lagutin et al., 2003; Martinez-Ferre & Martinez, 2012; Stevens et al., 2010), our scRNA-seq data revealed that the *ZIC2*<sup>-/-</sup> neural progenitors acquire a rostral telencephalic identity (i.e. high expression of *SIX3*, *LHX2*, *FOXP1*, *FGF8*) (Figure 4.17D) with very few cells expressing *SHH* (Figure 4.17C). Therefore, in the absence of high SHH levels, it is not surprising that ventral genes remained lowly expressed in *ZIC2*<sup>-/-</sup> neural progenitors. Nevertheless, we wondered whether the loss of WNT in *ZIC2*<sup>-/-</sup> neural progenitors, which already occurs on day 4 (see WNT1 in Figure 4.9), could affect their responsiveness to SHH signaling in comparison to WT cells. To test this, WT and *ZIC2*<sup>-/-</sup> hiPSC were differentiated into AntNPC with the addition of either DMSO or 0.5μM of the SHH signaling agonist SAG (Smoothed Agonist) from day 4 to day 8. Cells were collected on day 8 and a single-cell RNA-seq experiment was performed.

Focusing first on WT cells (control and SAG-treated cells), UMAP visualization showed that cells segregated according to their treatment status, thus indicating that the treatment with SAG led to significant transcriptional changes (Figure 4.18A). Indeed, the expression of major dorsal/cortical hem genes, including *ZIC2*, *WNT3A* or *LMX1A*, was reduced. Moreover, the expression of ventral identity genes considered as SHH target, such as *FOXA2*, *NKX6-1* and, to a lesser extent, *NKX2-1* and *NKX2-2*, were up-regulated in the SAG-treated cells (Figure 4.18B and C). Thus, SAG treatment during the differentiation into AntNPC successfully induced the expression of the ventral identity genes in WT cells, while reducing the levels of dorsal ones. However, it is worth noting that the most ventral genes, such as *NKX2-1* and *NKX2-2*, which are considered to be directly activated by SHH/GLI (i.e. Class II genes) and require high levels of SHH signaling

(Briscoe, 2009; Briscoe et al., 2000), were only induced in a few WT cells with the used SAG concentration (Figure 4.18C).



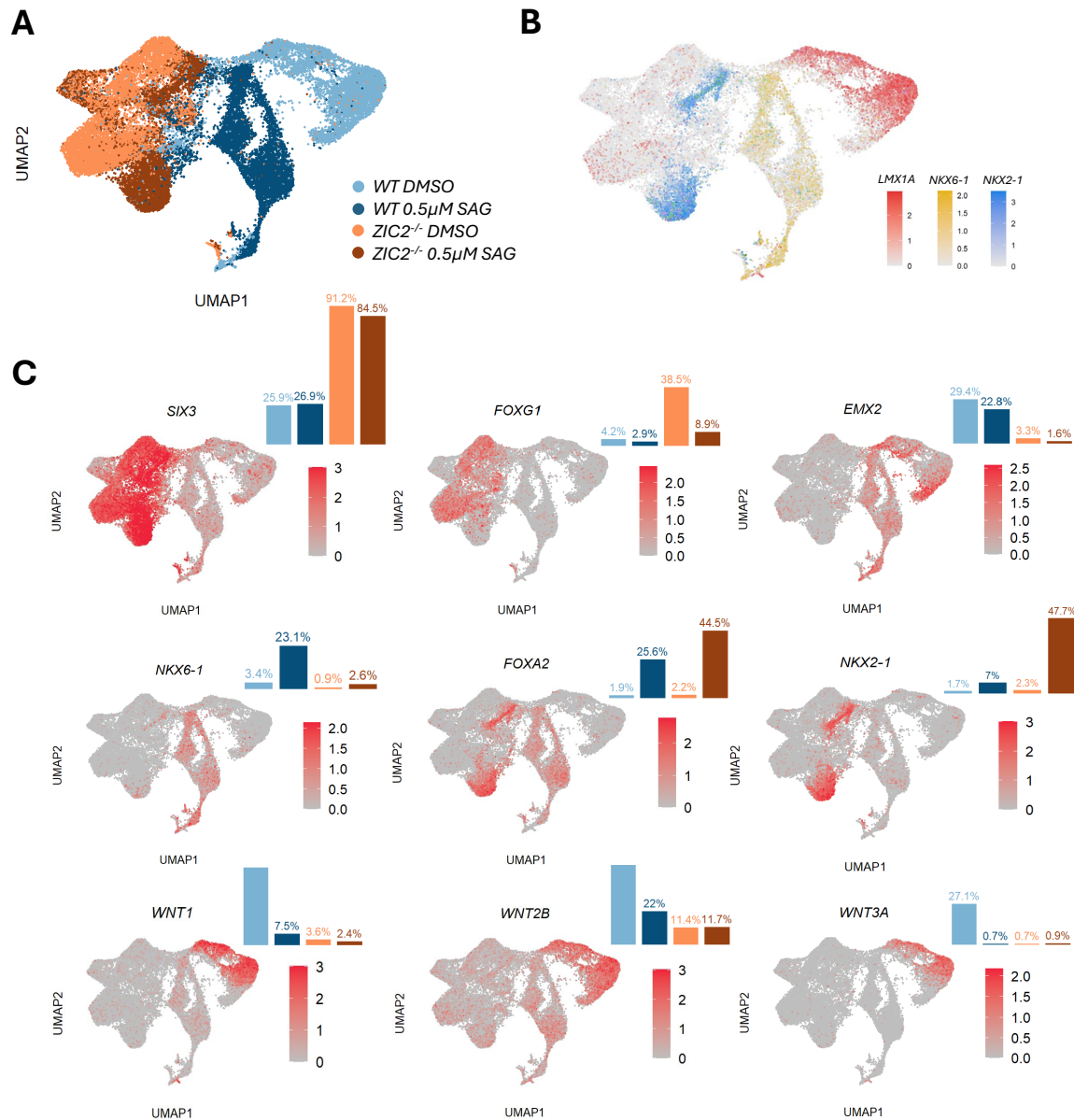
**Figure 4.18: SAG treatment in WT AntNPC leads to a loss of dorsal identities in favor of more ventral ones.**

(A) UMAP plot of WT neural progenitors on day 8, either untreated (light blue) or treated with 0.5 $\mu$ M of SAG (dark blue). (B) Dot plot representing the level of expression (color of the dot, z-score) and the percentage of cells (size of the dot) expressing the indicated genes in WT AntNPC, either untreated or treated with 0.5 $\mu$ M of SAG. (C) UMAP plots as in A showing the expression of neural patterning genes. Expression is calculated as  $\log_{10}(\text{feature counts per cell} / \text{total counts per cell}) \times 10000$ . Bar plots represent the percentage of cells expressing the gene in each sample.

Next, we visualized all four samples (WT untreated, WT SAG-treated, *ZIC2*<sup>-/-</sup> untreated and *ZIC2*<sup>-/-</sup> SAG-treated) in the UMAP plot to evaluate their response to the SAG treatment (Figure 4.19A). In the UMAP space, we plotted three major markers of the neural tube dorsal-ventral axis: *LMX1A* marks dorsal/cortical hem identities, *NKX6-1* marks intermediate progenitors (i.e. basal plate) and *NKX2-1* is expressed in the most ventral domain (Figure 4.19B). Notably, the scRNA-seq data revealed that *LMX1A* was mainly expressed in untreated WT cells, *NKX6-1* in WT cells treated with SAG and *NKX2-1* in *ZIC2*<sup>-/-</sup> cells treated with SAG. Thus, WT and *ZIC2*<sup>-/-</sup> cells responded differently to the SAG treatment: under the same dose of SAG, many WT cells induced *NKX6-1* (23.1%) but not *NKX2-1* (7%), while the opposite was observed for *ZIC2*<sup>-/-</sup> cells (2.6% of *NKX6-1* expressing cells; 47.7% of *NKX2-1* expressing cells). These results suggest that, in the absence of WNT, *ZIC2*<sup>-/-</sup> cells are hyper-responsive to SHH signaling and, under relatively low SHH/SAG levels, they induce the most ventral genes (i.e. Class II genes) (Briscoe, 2009; Briscoe et al., 2000). Similar results were obtained when considering other genes preferentially expressed in either dorsal or ventral domains: (i) genes similar to *NKX2-1* that are preferentially expressed in the most ventral domains (i.e. Class II genes; *FOXA2*, *NKX2-2*, *NKX2-8*, *GLI1*) were only induced upon SAG treatment, but the percentage of expressing cells was much higher in *ZIC2*<sup>-/-</sup> than in WT cells (Figure 4.19C, 19C); (ii) the expression of genes, which similarly to *LMX1A* are preferentially expressed in the most dorsal domains (cortical hem) (i.e. *WNT1*, *WNT2B* and *WNT3A*), was strongly reduced in WT cells upon SAG treatment and in both untreated and SAG treated *ZIC2*<sup>-/-</sup> cells (Figure 4.19C and 4.20A). Overall, these results suggest that, in the absence of WNT, *ZIC2*<sup>-/-</sup> cells are hyper-responsive to SHH signaling and, under relatively low SHH/SAG levels, they induce the most ventral genes (i.e. Class II genes) (Briscoe, 2009; Briscoe et al., 2000).

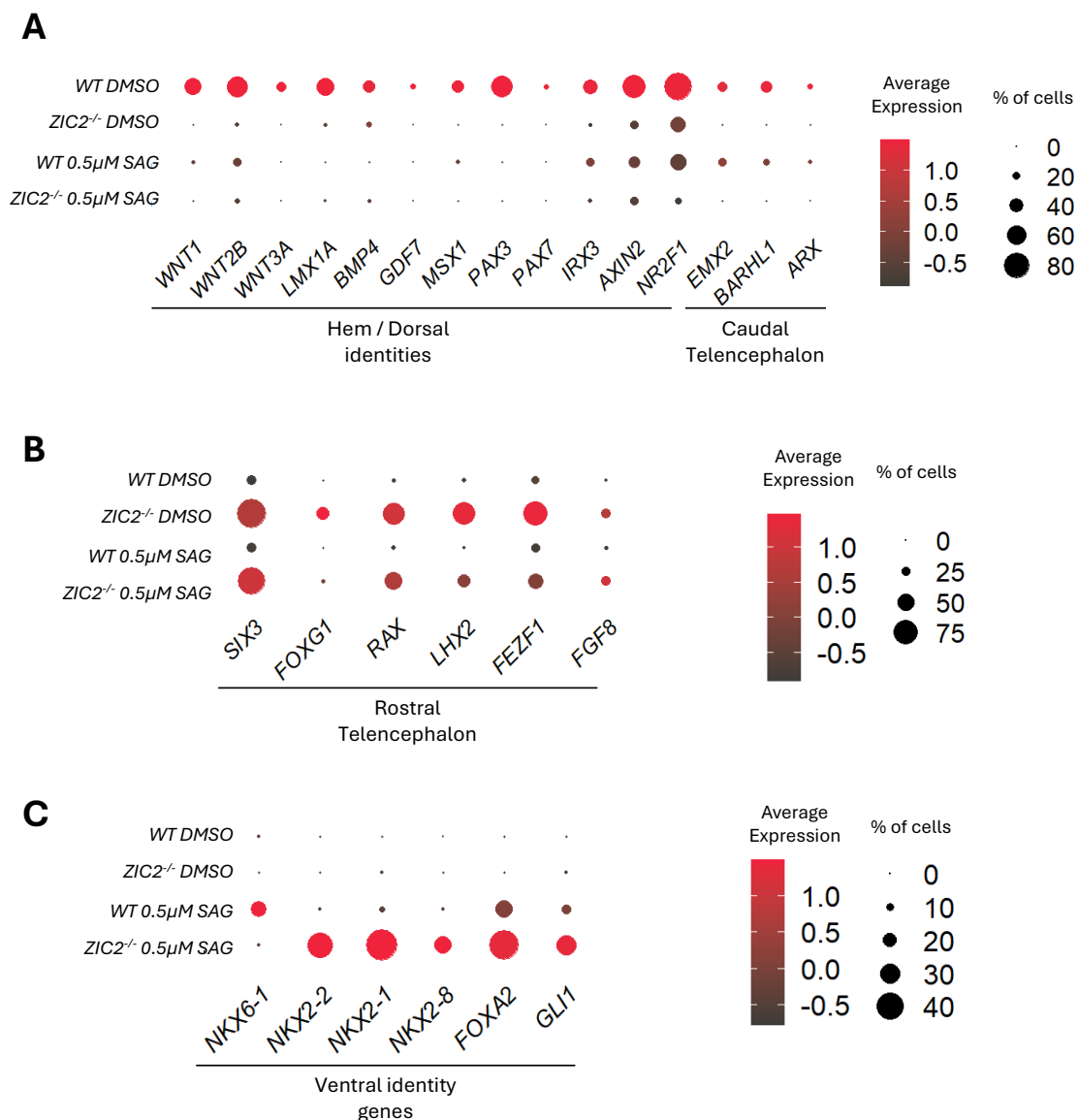
Since the signals controlling the dorsoventral patterning of the brain can also have an effect on the rostro-caudal axis, we also evaluated the expression of genes expressed in rostral telencephalon and whose expression was increased in *ZIC2*<sup>-/-</sup> cells (e.g. *SIX3*, *FOXG1*, *RAX*, *LHX2*) (Figure 4.19C, 19B). Interestingly, the number of *ZIC2*<sup>-/-</sup> cells expressing these rostral genes was in general reduced by the SAG treatment, in some cases (e.g. *FOXG1*) quite strongly (Figure

4.19C). These results indicate that, in *ZIC2*<sup>-/-</sup> cells, the SAG treatment not only induced the most ventral forebrain identity genes but also had a posteriorizing effect.



**Figure 4.19: *ZIC2*<sup>-/-</sup> neural progenitors are sensitized to SAG treatment.**

(A) UMAP plot of WT untreated (light blue), WT SAG-treated (dark blue), *ZIC2*<sup>-/-</sup> untreated (orange) and *ZIC2*<sup>-/-</sup> SAG-treated (brown) neural progenitors on day 8. (B) UMAP plot as in A. Expression of *LMX1A* (red), *NKX6-1* (yellow) and *NKX2-1* (blue) is highlighted. (C) UMAP plots of the expression of key target genes in WT and *ZIC2*<sup>-/-</sup> neural progenitors. Expression is calculated as  $\log_{10}(\text{feature counts per cell} / \text{total counts per cell}) \times 10000$ .



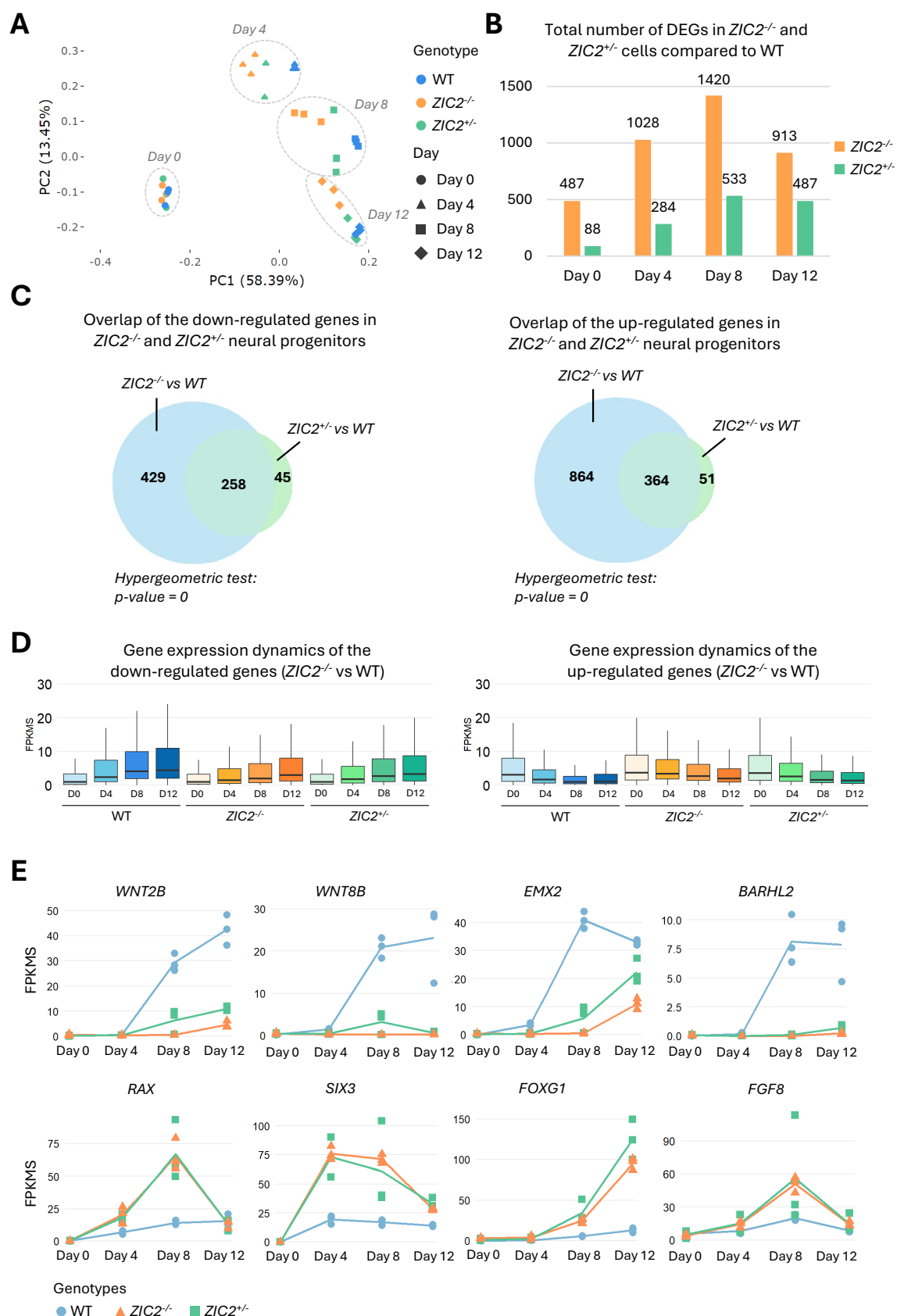
**Figure 4.20: Dot plot of expression of key neural patterning genes upon SAG treatment.**

Dot plots representing the level of expression (color of the dot, z-score) and the percentage of cells expressing (size of the dot) the indicated genes in WT and ZIC2<sup>-/-</sup> neural progenitors, untreated or treated with 0.5μM of SAG. (A) Markers of the dorsal forebrain and the caudal telencephalon. (B) Markers of the rostral telencephalon. (C) Markers of ventral identity genes that are SHH target genes: Class I (NKX6-1) and Class II (NKX2-1, NKX2-2, NXX2-8) genes.

#### 4.10. Reduced ZIC2 dosage also leads to significant transcriptional defects in AntNPCs

HPE cases caused by loss of *ZIC2* function typically present heterozygous mutations or deletions affecting this gene. Therefore, we finally decided to interrogate whether we could also detect *ZIC2* dosage sensitivity in our AntNPC differentiation system. In other words, we wanted to test whether the defects observed in a full knock-out of *ZIC2* would also persist, at least to some extent, in cells in which only one *ZIC2* allele is lost (i.e. 50% *ZIC2* dosage). For that purpose, we isolated three *ZIC2*<sup>+/-</sup> hiPSC clonal lines (Figure 4.3) that we differentiated into AntNPC in parallel with WT and *ZIC2*<sup>-/-</sup> hiPSC. RNA was collected throughout the differentiation (on day 0, day 4, day 8 and day 12) and analyzed by bulk RNA-seq as previously described for WT and *ZIC2*<sup>-/-</sup> cells. PCA showed that, on day 0, all samples clustered close to each other, indicating again that the loss of *ZIC2*, in either homozygosis or heterozygosis, does not have major impact on pluripotency. As described in previous sections (Figure 4.6), upon differentiation WT and *ZIC2*<sup>-/-</sup> cells separated from each other, indicating important transcriptional differences between the two genotypes. Furthermore, upon differentiation, the *ZIC2*<sup>+/-</sup> cells were located between the WT and *ZIC2*<sup>-/-</sup> samples, which was also reflected in a considerable number of DEGs with respect to WT cells, albeit lower than in *ZIC2*<sup>-/-</sup> cells (Figure 4.21B). Interestingly, 86% of the DEGs genes found in *ZIC2*<sup>+/-</sup> cells on day 4 and day 8 were also misregulated in *ZIC2*<sup>-/-</sup> neural progenitors. This significant overlap indicates that a 50% reduction in *ZIC2* dosage can lead to transcriptional defects similar to the ones found in *ZIC2*<sup>-/-</sup> cells (Figure 4.21C). This observation was further supported when evaluating the expression dynamics of all the up and down-regulated genes found in *ZIC2*<sup>-/-</sup> cells. As shown earlier (Figure 4.16D), down-regulated genes in *ZIC2*<sup>-/-</sup> cells mostly correspond to genes that become induced upon differentiation of hiPSC into AntNPC and that failed to do so in the absence of *ZIC2*. A similar, albeit more moderate, defect in gene inductions was observed in the *ZIC2* heterozygous cells. On the other hand, genes that were up-regulated in *ZIC2*<sup>-/-</sup> cells largely corresponded to genes expressed on day 0 that become silenced upon differentiation and that failed to be repressed in the absence of *ZIC2*. Once again, in the *ZIC2*<sup>+/-</sup> cell line, similar transcriptional defects were observed (Figure 4.21D). Notably, the expression dynamics of key *ZIC2* target genes that were described in previous sections and that play major roles in forebrain patterning (Figure 4.21E) (e.g. *WNT2B*, *WNT8B*, *EMX2*, *BARHL2*, *EMX2* and *BARHL2*) were similarly and severely compromised in both *ZIC2*<sup>-/-</sup> and *ZIC2*<sup>+/-</sup> cells. Thus, during AntNPC differentiation, neural progenitors are highly sensitive to *ZIC2* dosage, displaying strong

transcriptional defects in genes with major regulatory functions during both rostro-caudal and dorsal-ventral patterning of the forebrain.



**Figure 4.21: *ZIC2*<sup>+/-</sup> neural progenitors are sensitive to reduced dosage of *ZIC2*.**

(A) Principal component analysis of the RNA-seq samples generated in WT, *ZIC2*<sup>-/-</sup> and *ZIC2*<sup>+/-</sup> at the different time points of the differentiation (day 0, day 4, day 8 and day 12). (B) Barplot representing the total number of differentially expressed genes in *ZIC2*<sup>-/-</sup> (orange) and *ZIC2*<sup>+/-</sup> (green) compared to WT at the different timepoints of the differentiation. (C) Venn diagram representing the overlap between the genes down or up-regulated in *ZIC2*<sup>-/-</sup> and in *ZIC2*<sup>+/-</sup> cells (day 4 and 8 combined). (D) Boxplot of the expression in FPKMS of the down and up-regulated genes (*ZIC2*<sup>-/-</sup> vs WT, day 4 and 8 combined) across genotypes (WT, *ZIC2*<sup>-/-</sup> and *ZIC2*<sup>+/-</sup>). (E) Gene expression dynamics in FPKMS of top target genes in WT (light blue), *ZIC2*<sup>-/-</sup> (orange) and *ZIC2*<sup>+/-</sup> (green) cells.

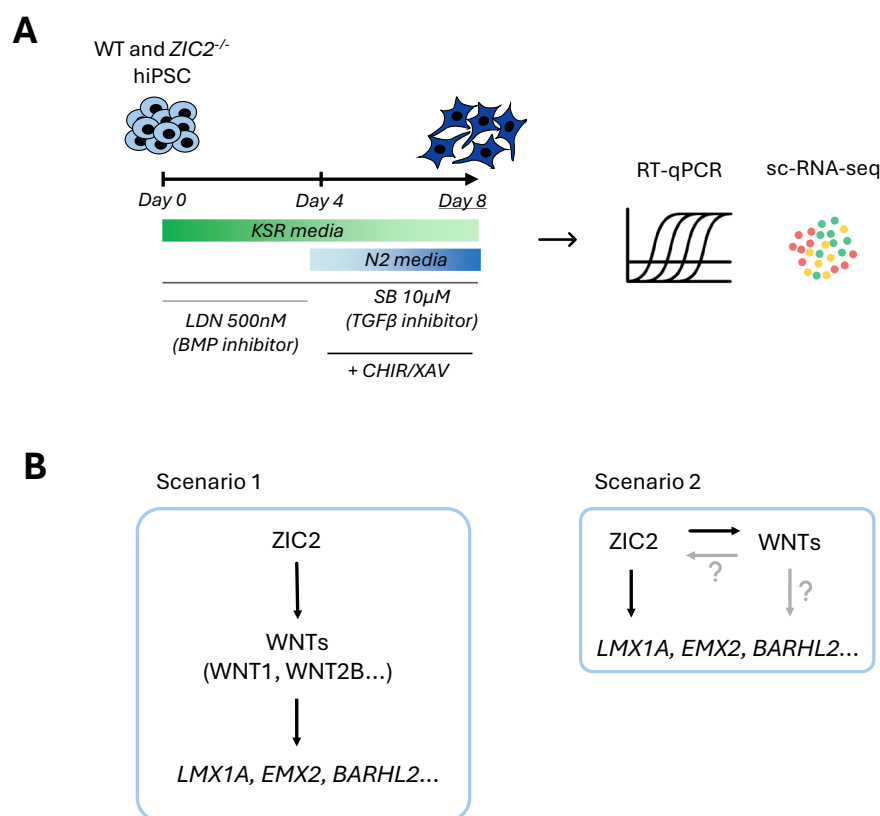
#### 4.11. Future experiments

In our differentiation system, one key target of *ZIC2* seems to be the WNT signaling pathway (i.e. strong down-regulation of WNT genes in *ZIC2*<sup>-/-</sup> AntNPC, binding of *ZIC2* in distal enhancers associated with WNT genes and losing acetylation in *ZIC2*<sup>-/-</sup> cells). To further study the interplay between *ZIC2* and WNT signaling, we are planning to differentiate WT and *ZIC2*<sup>-/-</sup> hiPSC into AntNPC with the addition of XAV (a WNT inhibitor) or CHIR (a WNT agonist) (Figure 4.22A) from day 4 to day 8. This experiment should help us distinguish between two possible scenarios (Figure 4.22B).

In the first one, *ZIC2* regulates the expression of WNT genes, which in turn regulate (via the canonical  $\beta$ -catenin pathway) the expression of important forebrain patterning genes promoting dorsal/cortical hem fates (e.g., *LMX1A*, *BMP4*, *GDF7*, *MSX1*, *PAX3*, *PAX7*) and preventing the upregulation of rostral telencephalon genes (Lagutin et al., 2003). Considering this scenario, CHIR treatment in *ZIC2*<sup>-/-</sup> neural progenitors should rescue the transcriptional defects observed upon the loss of *ZIC2* (i.e., up-regulation of dorsal/cortical hem and caudal telencephalon/diencephalon markers and down-regulation of the rostral telencephalon markers), whereas in WT cells, an increase in caudal telencephalon/diencephalon and a loss of rostral telencephalon markers should be observed. On the other hand, XAV treatment in WT cells should recapitulate to some extent the transcriptional defects originally observed in *ZIC2*<sup>-/-</sup> cells (i.e., downregulation of dorsal/cortical hem and caudal telencephalon/diencephalon markers in favor of the rostral-most identities). As WNT genes are already down-regulated in *ZIC2*<sup>-/-</sup> AntNPC, little to no changes should be observed upon XAV treatment.

In the second scenario, *ZIC2* and WNT genes would regulate each other in a positive feedback loop and *ZIC2*, perhaps together with  $\beta$ -catenin/TCF, would directly control the expression of the *ZIC2* main targets. Indeed, *ZIC2* seems to directly control enhancers that are not only associated with WNT genes but also with other important forebrain patterning regulators such as *LMX1A*,

*BARHL1*, *BARHL2* or *EMX2*. According to this model, CHIR treatment should rescue the transcriptional defects of the *ZIC2*<sup>-/-</sup> cells only partially, while XAV treatment in WT cells might have more profound consequences as it might decrease the expression of *ZIC2*. To further evaluate the coordinated action of *ZIC2* and WNT in the establishment of forebrain identities, ChIP-seq experiment of  $\beta$ -catenin, the nuclear effector of canonical WNT signaling, will be performed.

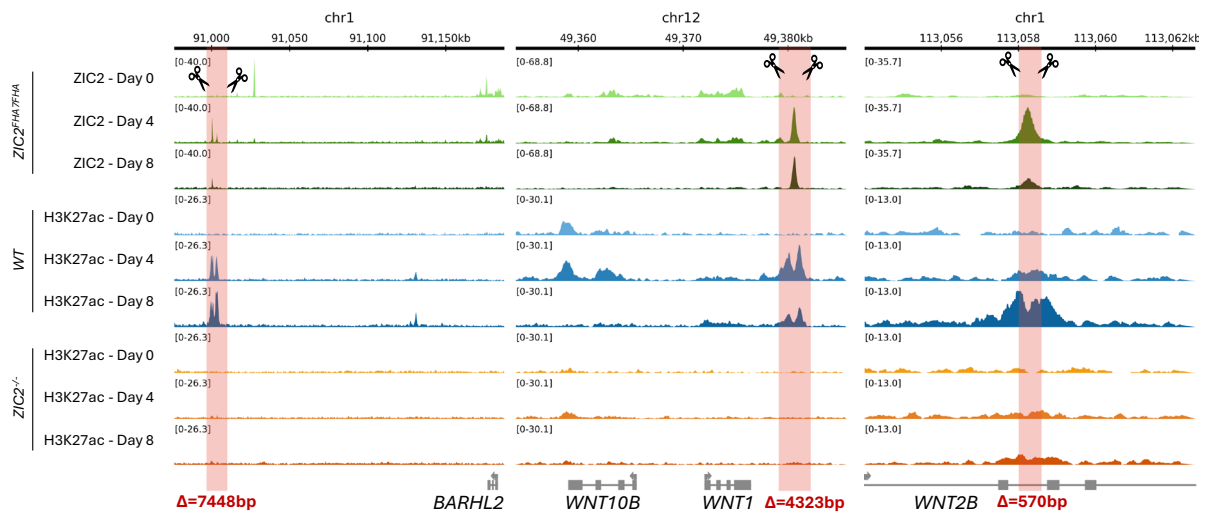


**Figure 4.22: Future experiments exploring the interplay between *ZIC2* and WNT signaling.**

(A) Schematic representation of the planned experiment. WT and *ZIC2*<sup>-/-</sup> hiPSC will be differentiated into AntNPC with the addition of CHIR or XAV. RNA will be collected on day 8 and expression of key target genes will be measured by RT-qPCR and scRNA-seq. (B) Diagram illustrating two different models whereby *ZIC2* and WNT signaling can control the expression of forebrain patterning genes.

Our results strongly depict *ZIC2* as an enhancer activator, regulating the expression of major anterior neural genes involved in forebrain patterning. To further confirm some of the predicted *ZIC2* targets, we decided to delete three putative enhancers located in proximity of *WNT1*, *WNT2B* and *BARHL2* genes (Figure 4.23). These enhancers were selected because they exhibited the following pattern: first, their predicted associated genes are expressed in WT neural progenitors but down-regulated upon *ZIC2* loss; second, these regions are bound by *ZIC2* at least on day 4 and/or day 8; third, these regions are marked with H3K27ac in WT cells, but this histone mark is

strongly reduced in *ZIC2*<sup>-/-</sup> cells. In the near future, using CRISPR-Cas9 technology, I will delete those predicted enhancers and derive clonal hiPSC with homozygous deletions for each of them. Next, the isolated hiPSC lines will be differentiated into AntNPC, in parallel with WT and *ZIC2*<sup>-/-</sup> controls. If the expression of the associated genes is downregulated in the corresponding cell line, it will further support the role of *ZIC2* as an activator of distal enhancers during the establishment of anterior neural identities.



**Figure 4.23: Deletion of predicted *ZIC2*-bound enhancers associated with expression of major anterior neural patterning genes.**

ChIP-seq tracks of *ZIC2* (*ZIC2*<sup>FHA/FHA</sup> line) and H3K27ac (WT and *ZIC2*<sup>-/-</sup>) on day 0, day 4 and day 8 are represented. Highlighted in red, the position of the three *ZIC2*-bound enhancers that will be deleted by CRISPR-Cas9. The sizes of the deletions are shown in red. Tracks figures were generated using Figeno (Sollier et al., 2024).

## 5. Discussion

In this project, we investigated the role of ZIC2 during human forebrain development and patterning using a hiPSC-based *in vitro* differentiation system. By combining an *in vitro* differentiation model of hiPSC into AntNPC with CRISPR-Cas9 genome editing and various genomic approaches, we were able to characterize the genes and cis-regulatory elements directly regulated by ZIC2 and that are sensitive to ZIC2 dosage during the establishment of forebrain-like neural identities. More specifically, our data shows that ZIC2 is an important enhancer activator during forebrain development and patterning, essential for the establishment of the gene expression programs associated with the dorsal forebrain/cortical hem (e.g., *WNT2B*, *WNT8B*, *LMX1A*) and the caudal telencephalon/rostral diencephalon (e.g., *EMX2*, *BARHL1*, *BARHL2*). Lastly, we also show that by promoting dorsal identities and WNT signaling, ZIC2 also modulates the responsiveness to hedgehog signaling in anterior neural progenitors, which might have important implications in the context of ZIC2-associated holoprosencephaly.

### 5.1. Modeling human forebrain development *in vitro*

To study the role of ZIC2 during neural induction, we used hiPSC and their differentiation towards AntNPC (Tchieu et al., 2017). This protocol describes the generation of neuroectoderm cells expressing the telencephalic markers *FOXG1* and *PAX6*. However, thanks to the generation of the single-cell transcriptomic profile of day 8 neural progenitor cells, we showed that the resulting neural progenitors have distinctive features of forebrain progenitors, recapitulating important aspects of dorsal-ventral and rostro-caudal patterning of the forebrain. Thus, this differentiation system allowed us to uncover key target genes of ZIC2. To further strengthen our findings, it would be interesting to use brain organoids derived from our hiPSC lines to model neural induction in three-dimensional space. Organoids have the advantages of modeling further aspects of brain complexity, including cellular interactions and three-dimensional morphology that better reflect *in vivo* cortex development (S. H. Kim & Chang, 2023). Moreover, several studies comparing single-cell RNA-seq data from human organoids with fetal brains showed that human organoids can faithfully recapitulate important aspects of *in vivo* cortex development and that human-specific gene expression patterns are conserved (Amiri et al., 2018; Camp et al., 2015; Kanton et al., 2019). Nevertheless, the use of brain organoids presents several limitations such as a higher batch-to-batch variability in comparison with 2D differentiation systems or longer experimental times. Furthermore, due to the high number of cellular identities generated within the organoids, they might not be suitable for bulk measurements, such as ChIP-seq, where cellular heterogeneity might confound the results. Another challenge when using brain organoids is the difficulty of directing the differentiation toward the appropriate cellular identities (i.e. cortical

hem), in order to properly study the function of *ZIC2* whose expression during brain development gets progressively restricted to the dorsal midline.

## 5.2. *ZIC2* in hiPSC

On day 0, few transcriptional changes are detected in *ZIC2*<sup>-/-</sup> hiPSC, along with few differences in H3K27ac levels genome-wide. Moreover, the induction of pan-neural markers (e.g. *SOX2*) is not impaired in *ZIC2*<sup>-/-</sup> cells. Therefore, the loss of the *ZIC2* gene in human pluripotent cells does not have major consequences. However, in hiPSC, expression of *ZIC2* is high and a large number of *ZIC2* binding events are detected. Why do these *ZIC2* binding events not have any obvious functional consequences? We foresee at least three scenarios that could explain the lack of any major defects in *ZIC2*<sup>-/-</sup> hiPSC:

- (i) Most of *ZIC2* binding sites on day 0 could be non-functional (Biggin, 2011; Spivakov, 2014), where promiscuous *ZIC2* binding occurs because the sites are accessible due to the particularly open chromatin and low DNA methylation levels that characterize pluripotent cells (Meshorer et al., 2006).
- (ii) Another possibility is that in pluripotent cells, the loss of *ZIC2* is compensated by other members of the *ZIC* family, in particular *ZIC3*, which is also expressed at high levels in hiPSC. This possibility is strongly supported by the partial functional redundancy of *ZIC2* and *ZIC3* during mouse development (Inoue et al., 2007) and in human pluripotency (Hossain et al., 2024) and their strongly overlapping binding profiles in mESC (Mariner-Fauli et al., manuscript in preparation).
- (iii) Lastly, *ZIC2* could act as a pioneer factor on day 0, preparing the chromatin for future activation. This hypothesis is supported by studies performed in mouse and human embryonic stem cells (Hossain et al., 2024; Luo et al., 2015; Mariner-Fauli et al., manuscript in preparation), where *ZIC2* has been reported to interact with chromatin remodelling complexes, such as SWI/SNF.

## 5.3. Dorsal-ventral patterning of forebrain-like is impaired in *ZIC2*<sup>-/-</sup> neural progenitors

Analyses of the bulk and single-cell RNA-seq data during AntNPC differentiation revealed that both dorsal-ventral and rostro-caudal patterning are affected in *ZIC2*<sup>-/-</sup> neural progenitors compared to their WT counterparts. More precisely, major cortical hem regulators were strongly downregulated (e.g., *WNT2B*, *WNT8B*, *LMX1A*) (Caronia-Brown et al., 2014; Iskusnykh et al., 2023) and a shift from caudal telencephalon/diencephalon identities towards more rostral ones

was observed (i.e., down-regulation of *EMX2* and *BARHL2* (diencephalon/caudal telencephalon) (Parish et al., 2016; Suda et al., 2001) and up-regulation of *SIX3*, *FOXP1*, *FGF8* (rostral telencephalon) (Carlin et al., 2012; Lagutin et al., 2003; Sato et al., 2017)). These results are consistent with the HPE phenotype associated with *ZIC2* loss of function both in mouse and human. Indeed, HPE results from an incomplete midline cleavage of the forebrain and impaired dorsal-ventral patterning. More specifically, the severe reduction in cortical hem identities and WNT signaling observed in *ZIC2*<sup>-/-</sup> and *ZIC2*<sup>+/-</sup> AntNPC is in perfect agreement with the unique involvement of *ZIC2* loss of function mutations in the middle interhemispheric variant (MIHV) of HPE in humans (Gounongbé et al., 2020) as well as with the neural tube closure defects and delayed differentiation of dorsal neural structures observed in *Zic2* mouse hypomorphs (Nagai et al., 2000). In contrast with the classical forms of HPE, which arise due to defects in the forebrain floor plate, the MIHV is considered to result from a defect in the induction of the cortical hem/roof plate (Gounongbé et al., 2020). Therefore, our results provide important insights into the molecular basis of how *ZIC2* haploinsufficiency can lead to the MIHV of HPE.

Among the defects observed in *ZIC2*<sup>-/-</sup> and *ZIC2*<sup>+/-</sup> AntNPC, the severe reduction in the expression of several WNT genes (e.g. *WNT2B*, *WNT3A*, *WNT1*) that are considered major cortical hem/roof plate markers was particularly interesting. We showed that *ZIC2* binds to several putative enhancers associated with these downregulated WNT genes and that showed reduced H3K27ac levels in *ZIC2*<sup>-/-</sup> neural progenitors. Importantly, previous studies suggest that WNT signaling represents a major and evolutionary conserved target of *ZIC2* both during neural development as well as in other cellular contexts. Firstly, *ZIC* orthologs in *Drosophila* and *Xenopus* can induce the expression of WNT ligands (Benedyk et al., 1993; Merzdorf & Sive, 2006), which is consistent with our findings. In the mouse developing retina, Morenilla-Palao et al. showed that *ZIC2* interacts with the components of the WNT signaling pathway. More precisely, they showed that the differential response of contralateral and ipsilateral axon projections to the WNT5A ligand can be explained by the expression of *ZIC2* in iRGCs but not in cRGCs, with *ZIC2* expression inducing different WNT receptors and, thus, triggering a different response to WNT5A (Morenilla-Palao et al., 2020). During neural crest specification, *ZICs* were reported to act as co-repressors of the WNT pathway in low WNT activity regions (i.e., lateral neuroectoderm), while having a more classical activator role in a high WNT context (i.e., neural plate border). The switch from co-repressor activity to activator could be favored by the SUMOylation of the *ZIC* proteins (Ali et al., 2021; Bellchambers et al., 2021). Although the interplay between *ZIC2* and WNT signaling has not been extensively addressed in the context of early forebrain development and patterning, previous work by the Aruga lab showed that *WNT3A* expression was reduced in the roof plate of

the developing telencephalon and spinal cord of *ZIC2* mouse hypomorphs (Nagai et al., 2000). Altogether, these studies strongly support the interaction between *ZIC2* and WNT signaling. The future experiments described in the result section (Figure 4.22), together with  $\beta$ -catenin ChIP-seq experiments during AntNPC differentiation, should help us further characterize the potential crosstalk between *ZIC2* and WNT signaling in the context of human forebrain development.

Several components of the SHH signaling pathway (e.g. *SHH*, *PTCH1*, *GLI2*) are often mutated in classical HPE (Dubourg et al., 2007; Tekendo-Ngongang et al., 1993), which results from defects in forebrain floor plate development. Furthermore, the WNT and SHH signaling pathways antagonize each other, promoting dorsal and ventral forebrain identities, respectively (Ulloa & Martí, 2010). Since we did not observe an up-regulation of ventral identity genes in *ZIC2*<sup>-/-</sup> neural progenitors despite the loss of WNT signaling and cortical hem/roof plate identities (Figure 4.9), we decided to investigate how the loss of *ZIC2* in neural progenitors could affect the responsiveness to hedgehog signaling, as this could provide additional insights into the mechanisms of *ZIC2*-associated HPE, which uniquely includes both the classical and the MIH variants. To do so, upon neural induction (i.e. day 4), we treated the AntNPC with SAG, an agonist of the hedgehog signaling pathway, and collected the cells on day 8 to perform scRNA-seq experiments. We demonstrated that in the absence of *ZIC2*, neural progenitors are sensitized to SAG treatment. Indeed, the same dose of SAG led to a much higher proportion of cells expressing class II genes representing direct SHH targets (i.e., *FOXA2*, *NKX2-2*, *NKX2-1*, *NKX2-8*, *GLI1*) in *ZIC2*<sup>-/-</sup> neural progenitors compared to WT, thus indicating that, in the absence of *ZIC2*, AntNPC acquired a more ventral identity. In addition, we also showed that the ventralization of *ZIC2*<sup>-/-</sup> neural progenitors upon SAG treatment is also accompanied by a milder posteriorizing effect, illustrated by the downregulation of rostral telencephalon markers such as *FOXP1* and *FGF8*.

Taking together, we speculate that, during forebrain patterning, *ZIC2* and WNT signaling could potentially play a similar role as GLIs and SHH signaling, but in a opposite concentration gradient. Namely, we hypothesize that in a low WNT signaling context (i.e. ventral neural progenitors), *ZIC2* might act as a (co)repressor (Ali et al., 2021; Bellchambers et al., 2021), analogous to *GLI3R* in low SHH contexts (i.e. dorsal neural progenitors) (Oosterveen et al., 2012). In contrast, high WNT activity would favor a role as a transcriptional activator for *ZIC2*, much like *GLI1* and *GLI2* inducing the expression of the class II genes in response to strong hedgehog signaling (Briscoe et al., 2000; Corbin et al., 2003; Fuccillo et al., 2004).

Last but not least, the role of *ZIC2* as an activator of major cortical hem markers and WNT ligand genes (e.g. *LMX1A*, *WNT3A*, *WNT2B*) helps understanding the molecular basis of the MIHV of

HPE, which, among the HPE-associated genes, is uniquely caused by mutations affecting *ZIC2*. However, *ZIC2* haploinsufficiency can also lead to the classical form of HPE, which is considered to be caused by floor plate defects and is often associated with mutations affecting SHH pathway genes (e.g. *SHH*, *PTCH1*, *GLI1*). How can *ZIC2* haploinsufficiency lead to either classical or MIHV HPE in different patients? Notably, previous studies have demonstrated that both gain or loss of activity of the SHH pathway can lead to human congenital limb defects (Lopez-Rios, 2016). Our data shows that in the absence of *ZIC2*, AntNPCs show a loss of cortical hem/roof plate identities and WNT signaling, which in turn sensitizes them to SHH signaling. We speculate that, similarly to putative SHH gain-of-function mutations identified in classical HPE patients, a reduction in *ZIC2* dosage might exacerbate SHH signaling in forebrain neural progenitors, which in turn can disrupt proper ventral patterning and lead to classical HPE (Casillas & Roelink, 2018). Under this model, whether *ZIC2* haploinsufficiency leads to either the classical or MIH variants of HPE would depend on additional interactions with genetic and/or environmental risk factors that could also influence forebrain dorsal-ventral patterning (e.g. alcohol consumption, pregestational diabetes) (Addissie et al., 2021; Lo et al., 2021).

#### **5.4. *ZIC2* as an enhancer binding factor**

Regarding the mechanism of action of *ZIC2*, our data shows that, upon neural induction (i.e. from day 4 onwards), this TF is acting as a transcriptional activator of distal enhancers that are essential for the proper expression of major forebrain patterning regulators (Figure 4.15 and 4.16). To further link *ZIC2* target genes with those distal enhancers bound by *ZIC2*, we will delete these putative enhancers using CRISPR-Cas9 technology in hiPSC. The resulting hiPSC clonal lines (homozygous for the deletions) will be differentiated into AntNPC, in parallel with WT and *ZIC2*<sup>-/-</sup> controls, to assess whether the expression of the putative target genes is affected. However, smaller deletions or point mutations restricted to the *ZIC2* binding sites would be necessary to further confirm *ZIC2* as an activator of those enhancers. Moreover, our data also suggests, albeit not as strongly, that upon pluripotency exit (i.e. from day 0 to day 4), *ZIC2* might act as a repressor of enhancers that are active in hiPSC and that become silenced as cells start to differentiate (Figure 4.13 and 4.14). Interestingly, those *ZIC2*-bound pluripotency enhancers were not enriched in the *ZIC2* motif (Figure 4.13B), suggesting that *ZIC2* might act as a co-repressor that gets recruited to these enhancers by other TFs. Such co-repressor role for *ZIC2* has been previously proposed in cellular contexts with low WNT activity (Ali et al., 2021; Bellchambers et al., 2021) and, accordingly, our RNA-seq expression data shows that the expression of WNT ligand genes is low in hiPSC (day 0) and starts increasing from day 4 onwards.

### 5.5. Human and mouse show different sensitivity to ZIC2 dosage during forebrain development

While generating *ZIC2*<sup>-/-</sup> hiPSC lines, we also isolated clones with heterozygous deletions of the *ZIC2* gene (i.e., *ZIC2*<sup>+/-</sup> hiPSC lines) in order to test whether the ZIC2 dosage sensitivity typically observed in ZIC2-related HPE patients would also be detected in our differentiation system. In parallel with the WT and *ZIC2*<sup>-/-</sup> hiPSC, the *ZIC2*<sup>+/-</sup> hiPSC lines were also differentiated into AntNPC and analyzed by bulk RNA-seq. These analyses revealed that the transcription defects previously described in *ZIC2*<sup>-/-</sup> neural progenitors were, to a large extent, also observed in *ZIC2*<sup>+/-</sup> cells (Figure 4.21). Interestingly, RNA-seq data that I previously generated in mouse WT, *Zic2*<sup>-/-</sup> and *Zic2*<sup>+/-</sup> ES cells and neural progenitor cells (Mariner-Fauli et al., manuscript in preparation) revealed that, contrary to my observations in human cells, the *Zic2*<sup>+/-</sup> neural progenitors barely showed any transcriptional defects in comparison to WT cells, while such defects were very pronounced in *Zic2*<sup>-/-</sup> neural progenitors in which, again, the expression of several brain patterning regulators was affected. These findings are consistent with the fact that heterozygous mice for *Zic2* do not show major forebrain abnormalities. Although the direct comparison of the two species should be done cautiously, these observations strongly suggest that mice and humans differ in their sensitivity to ZIC2 dosage during brain development and raise questions regarding the molecular basis (e.g. differences in the regulatory network/mechanism of ZIC2 between the two species) of such differences that might be worth investigating in the future.

## 6. Conclusions

In this thesis we investigated the role of the transcription factor ZIC2 upon anterior neural induction in human cells and made the following conclusions:

1. The differentiation of hiPSC into AntNPC successfully recapitulates important aspects of the dorsal-ventral and rostro-caudal patterning that occur during early forebrain development.
2. Using CRISPR-Cas9 technology, several hiPSC lines were generated, namely *ZIC2*<sup>+/+</sup>, *ZIC2*<sup>-/-</sup> and *ZIC2*<sup>FHA/FHA</sup>, which allowed us to interrogate ZIC2 function during the differentiation towards AntNPC.
3. In pluripotency, no major transcriptional defects are observed upon loss of the *ZIC2* gene and the capacity of *ZIC2*<sup>-/-</sup> hiPSC lines to differentiate towards anterior neural identities is not impaired.
4. Additionally, differential binding analysis of H3K27ac in *ZIC2*<sup>-/-</sup> vs WT hiPSC revealed that only a few regions show significant differences in H3K27ac, which is in agreement with the minor transcriptional changes observed in *ZIC2*<sup>-/-</sup> hiPSC.
5. Differential expression analysis of *ZIC2*<sup>-/-</sup> vs WT neural progenitors on day 4, day 8 and day 12 revealed that the highest number of DEGs occurs on day 8 and that 80% of the DEGs are already discovered by day 8.
6. Upon loss of the ZIC2 gene in neural progenitor cells, a shift in positional identity along the rostro-caudal axis is observed. Specifically, expression of anterior telencephalon markers (e.g., *SIX3*, *FOXP1*) is increased at the expense of the identities marking the posterior telencephalon and diencephalon (e.g., *EMX2*, *BARHL1*, *BARHL2*).
7. Dorsal-ventral patterning is also affected by the loss of ZIC2, as *ZIC2*<sup>-/-</sup> neural progenitors show reduced levels of WNT signaling genes (e.g., *WNT1*, *WNT2B*, *WNT8B*) and other markers of the cortical hem (e.g., *LMX1A*, *MSX1*, *GDF7*).
8. Single-cell RNA-seq experiments confirmed that, upon loss of the *ZIC2* gene, the proportion of cells expressing anterior telencephalon markers is increased and that the most dorsal identities are lost.
9. ChIP-seq of ZIC2 revealed that on day 0 and day 4, ZIC2 binds extensively throughout the genome, with 28% of *de novo* binding events occurring on day 4. The number of binding events is strongly reduced on day 8.

10. ZIC2 is able to bind both proximal (33%) and distal (66%) regions of the genome, that are either in an active (H3K27ac) or poised/inactive state (H3K27me3).
11. Analysis of the ZIC2-bound regions gaining H3K27ac in *ZIC2*<sup>-/-</sup> cells revealed that ZIC2 binds to CREs that are active in hiPSC and failed to be repressed upon *ZIC2* deletion. These regions were correlated with genes whose expression failed to be silenced upon neural induction in the absence of ZIC2. However, these regions are not enriched in the ZIC2 motif, suggesting that the repressing role of ZIC2 might be indirect, with ZIC potentially acting as a co-repressor recruited by other TFs.
12. Analysis of the ZIC2-bound regions losing H3K27ac in *ZIC2*<sup>-/-</sup> cells revealed that ZIC2 binds to distal CREs (i.e., enhancers) that become activated upon neural induction (day 4 and/or day 8) and that are associated with genes involved in forebrain development and patterning whose induction is compromised in the absence of ZIC2. These distal CREs are enriched in ZIC2 binding motif, arguing for a more direct role of ZIC2 as an activator of these enhancers.
13. Responsiveness to hedgehog signaling in *ZIC2*<sup>-/-</sup> neural progenitors is increased, with a higher number of cells expressing the direct targets of the SHH pathway (i.e., class II genes) in *ZIC2*<sup>-/-</sup> compared to WT cells upon exposure to similar levels of a SHH agonist (SAG).
14. Transcriptional analysis of *ZIC2*<sup>+/-</sup> neural progenitors showed that *ZIC2*<sup>+/-</sup> AntNPC are sensitive to ZIC2 dosage, displaying strong transcriptional defects similar to the ones observed in *ZIC2*<sup>-/-</sup> AntNPC.
15. The role of ZIC2 as an activator of major cortical hem markers and WNT ligand genes expressed in the dorsal midline of the forebrain helps in understanding the molecular basis of the MIVH form of HPE, which is characterized by dorsal midline defects.
16. Classical forms of HPE are characterized by ventral defects. We have shown that upon the loss of ZIC2, AntNPCs are sensitized to SHH signaling, which promotes ventral identities. This finding could explain why patients with ZIC2 haploinsufficiency can present both, MIVH (dorsal midline defects) and classical (ventral midline defects) forms of HPE.

## References

- Addissie, Y. A., Troia, A., Wong, Z. C., Everson, J. L., Kozel, B. A., Muenke, M., Lipinski, R. J., Malecki, K. M. C., & Kruszka, P. (2021). Identifying environmental risk factors and gene–environment interactions in holoprosencephaly. *Birth Defects Research*, 113(1), 63–76. <https://doi.org/10.1002/bdr2.1834>
- Ali, R. G., Bellchambers, H. M., Warr, N., Ahmed, J. N., Barratt, K. S., Neill, K., Diamand, K. E. M., & Arkell, R. M. (2021). WNT-responsive SUMOylation of ZIC5 promotes murine neural crest cell development, having multiple effects on transcription. *Journal of Cell Science*, 134(9). <https://doi.org/10.1242/jcs.256792>
- Amiri, A., Coppola, G., Scuderi, S., Wu, F., Roychowdhury, T., Liu, F., Pochareddy, S., Shin, Y., Safi, A., Song, L., Zhu, Y., Sousa, A. M. M., Gerstein, M., Crawford, G. E., Sestan, N., Abyzov, A., & Vaccarino, F. M. (2018). Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science*, 362(6420). <https://doi.org/10.1126/science.aat6720>
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Balaskas, N., Ribeiro, A., Panovska, J., Dessaud, E., Sasai, N., Page, K. M., Briscoe, J., & Ribes, V. (2012). Gene regulatory logic for reading the sonic hedgehog signaling gradient in the vertebrate neural tube. *Cell*, 148(1–2), 273–284. <https://doi.org/10.1016/j.cell.2011.10.047>
- Barratt, K. S., & Arkell, R. M. (2018). ZIC2 in holoprosencephaly. In *Advances in Experimental Medicine and Biology* (Vol. 1046). [https://doi.org/10.1007/978-981-10-7311-3\\_14](https://doi.org/10.1007/978-981-10-7311-3_14)
- Bellchambers, H. M., Barratt, K. S., Diamand, K. E. M., & Arkell, R. M. (2021). SUMOylation potentiates ZIC protein activity to influence murine neural crest cell specification. *International Journal of Molecular Sciences*, 22(19). <https://doi.org/10.3390/ijms221910437>
- Benedyk, M. J., Mullen, J. R., & Dinardo, S. (1993). *odd-paired: a zinc finger pair-rule protein required for the timely activation of engrailed and wingless in Drosophila embryos*.
- Biggin, M. D. (2011). Animal Transcription Networks as Highly Connected, Quantitative Continua. In *Developmental Cell* (Vol. 21, Issue 4, pp. 611–626). <https://doi.org/10.1016/j.devcel.2011.09.008>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., Xu, X., Lv, X., Hugnot, J. P., Tanay, A., & Cavalli, G. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*, 171(3), 557–572.e24. <https://doi.org/10.1016/j.cell.2017.09.043>
- Braun, E., Danan-Gotthold, M., Borm, L. E., Lee, K. W., Vinsland, E., Lönnerberg, P., Hu, L., Li, X., He, X., Andrusivová, Ž., Lundberg, J., Barker, R. A., Arenas, E., Sundström, E., &

- Linnarsson, S. (2023). Comprehensive cell atlas of the first-trimester developing human brain. *Science*, 382(6667). <https://doi.org/10.1126/science.adf1226>
- Breedlove, S. M. (2017). *FOUNDATIONS OF NEURAL DEVELOPMENT*.
- Briscoe, J. (2009). Making a grade: Sonic Hedgehog signalling and the control of neural cell fate. In *EMBO Journal* (Vol. 28, Issue 5, pp. 457–465). <https://doi.org/10.1038/emboj.2009.12>
- Briscoe, J., Pierani, A., Jessell, T. M., & Ericson, J. (2000). A Homeodomain Protein Code Specifies Progenitor Cell Identity and Neuronal Fate in the Ventral Neural Tube. In *Cell* (Vol. 101).
- Briscoe, J., & Small, S. (2015). Morphogen rules: design principles of gradient-mediated embryo patterning. *Development*, 142(23), 3996–4009. <https://doi.org/10.1242/dev.129452>
- Brown, L., & Brown, S. (2009). Zic2 is expressed in pluripotent cells in the blastocyst and adult brain expression overlaps with makers of neurogenesis. *Gene Expression Patterns*, 9(1), 43–49. <https://doi.org/10.1016/j.gep.2008.08.002>
- Bulchand, S., Grove, E. A., Porter, F. D., & Tole, S. (2000). *LIM-homeodomain gene Lhx2 regulates the formation of the cortical hem*. [www.elsevier.com/locate/modo](http://www.elsevier.com/locate/modo)
- Bulfone, A., Menguzzato, E., Broccoli, V., Marchitello, A., Gattuso, C., Mariani, M., Giacomo Consalez, G., Martinez, S., Ballabio, A., & Banfi, S. (2000). Barhl1, a gene belonging to a new subfamily of mammalian homeobox genes, is expressed in migrating neurons of the CNS. In *Human Molecular Genetics* (Vol. 9, Issue 9).
- Camp, J. G., Badsha, F., Florio, M., Kanton, S., Gerber, T., Wilsch-Bräuninger, M., Lewitus, E., Sykes, A., Hevers, W., Lancaster, M., Knoblich, J. A., Lachmann, R., Pääbo, S., Huttner, W. B., & Treutlein, B. (2015). Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proceedings of the National Academy of Sciences of the United States of America*, 112(51), 15672–15677. <https://doi.org/10.1073/pnas.1520760112>
- Carlin, D., Sepich, D., Grover, V. K., Cooper, M. K., Solnica-Krezel, L., & Inbal, A. (2012). Six3 cooperates with hedgehog signaling to specify ventral telencephalon by promoting early expression of foxg1a and repressing Wnt signaling. *Development (Cambridge)*, 139(14), 2614–2624. <https://doi.org/10.1242/dev.076018>
- Caronia-Brown, G., Yoshida, M., Gulden, F., Assimacopoulos, S., & Grove, E. A. (2014). The cortical hem regulates the size and patterning of neocortex. *Development (Cambridge)*, 141(14), 2855–2865. <https://doi.org/10.1242/dev.106914>
- Casillas, C., & Roelink, H. (2018). Gain-of-function Shh mutants activate Smo cell-autonomously independent of Ptch1/2 function. *Mechanisms of Development*, 153, 30–41. <https://doi.org/10.1016/j.mod.2018.08.009>
- Chambers, S. M., Fasano, C. A., Papapetrou, E. P., Tomishima, M., Sadelain, M., & Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnology*, 27(3), 275–280. <https://doi.org/10.1038/nbt.1529>
- Chen, X., Huang, Y., Huang, L., Huang, Z., Hao, Z. Z., Xu, L., Xu, N., Li, Z., Mou, Y., Ye, M., You, R., Zhang, X., Liu, S., & Miao, Z. (2024). A brain cell atlas integrating single-cell transcriptomes

- across human brain regions. *Nature Medicine*. <https://doi.org/10.1038/s41591-024-03150-z>
- Chen, Y., Chen, L., Lun, A. T. L., Baldoni, P. L., & Smyth, G. K. (2025). edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Research*, 53(2). <https://doi.org/10.1093/nar/gkaf018>
- Chizhikov, V. V., & Iskusnykh, I. Y. (2025). Cortical hem signaling center: functions, development, and potential implications for evolution and brain disorders. *Neural Regeneration Research*, 20(4), 1079–1080. <https://doi.org/10.4103/nrr.nrr-d-23-01796>
- Claringbould, A., & Zaugg, J. B. (2021). Enhancers in disease: molecular basis and emerging treatment strategies. In *Trends in Molecular Medicine* (Vol. 27, Issue 11, pp. 1060–1073). Elsevier Ltd. <https://doi.org/10.1016/j.molmed.2021.07.012>
- Corbin, J. G., Rutlin, M., Gaiano, N., & Fishell, G. (2003). Combinatorial function of the homeodomain proteins Nkx2.1 and Gsh2 in ventral telencephalic patterning. *Development*, 130(20), 4895–4906. <https://doi.org/10.1242/dev.00717>
- Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., van Ijcken, W. F. J., Grosveld, F., Frommolt, P., Bazzi, H., & Rada-Iglesias, A. (2017). PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell*, 20(5), 689–705.e9. <https://doi.org/10.1016/j.stem.2017.02.004>
- Delás, M. J., & Briscoe, J. (2020). Repressive interactions in gene regulatory networks: When you have no other choice. In *Current Topics in Developmental Biology* (Vol. 139, pp. 239–266). Academic Press Inc. <https://doi.org/10.1016/bs.ctdb.2020.03.003>
- Diamand, K. E. M., Barratt, K. S., & Arkell, R. M. (2018). *Overview of Rodent Zic Genes* (pp. 179–207). [https://doi.org/10.1007/978-981-10-7311-3\\_10](https://doi.org/10.1007/978-981-10-7311-3_10)
- Dubourg, C., Bendavid, C., Pasquier, L., Henry, C., Odent, S., & David, V. (2007). Holoprosencephaly. *Orphanet Journal of Rare Diseases*, 2(1), 8. <https://doi.org/10.1186/1750-1172-2-8>
- Elms, P., Scurry, A., Davies, J., Willoughby, C., Hacker, T., Bogani, D., & Arkell, R. (2004). Overlapping and distinct expression domains of Zic2 and Zic3 during mouse gastrulation. *Gene Expression Patterns*, 4(5), 505–511. <https://doi.org/10.1016/j.modgep.2004.03.003>
- Feng, J., Liu, T., Qin, B., Zhang, Y., & Liu, X. S. (2012). Identifying ChIP-seq enrichment using MACS. *Nature Protocols*, 7(9), 1728–1740. <https://doi.org/10.1038/nprot.2012.101>
- Fernandes, M., Gutin, G., Alcorn, H., McConnell, S. K., & Hébert, J. M. (2007). Mutations in the BMP pathway in mice support the existence of two molecular classes of holoprosencephaly. *Development*, 134(21), 3789–3794. <https://doi.org/10.1242/dev.004325>
- Fuccillo, M., Rallu, M., McMahon, A. P., & Fishell, G. (2004). Temporal requirement for hedgehog signaling in ventral telencephalic patterning. *Development*, 131(20), 5031–5040. <https://doi.org/10.1242/dev.01349>

- Fuccillo, M., Rutlin, M., & Fishell, G. (2006). Removal of Pax6 partially rescues the loss of ventral structures in Shh null mice. *Cerebral Cortex*, 16(SUPPL. 1).  
<https://doi.org/10.1093/cercor/bhk023>
- Furushima, K., Murata, T., Matsuo, I., & Aizawa, S. (2000). A new murine zinc finger gene, *Opr*.  
[www.elsevier.com/locate/mode](http://www.elsevier.com/locate/mode)
- Geng, X., Acosta, S., Lagutin, O., Gil, H. J., & Oliver, G. (2016). Six3 dosage mediates the pathogenesis of holoprosencephaly. *Development (Cambridge)*, 143(23), 4462–4473.  
<https://doi.org/10.1242/dev.132142>
- Geng, X., Speirs, C., Lagutin, O., Inbal, A., Liu, W., Solnica-Krezel, L., Jeong, Y., Epstein, D. J., & Oliver, G. (2008). Haploinsufficiency of Six3 Fails to Activate Sonic hedgehog Expression in the Ventral Forebrain and Causes Holoprosencephaly. *Developmental Cell*, 15(2), 236–247. <https://doi.org/10.1016/j.devcel.2008.07.003>
- Gounongbé, C., Marangoni, M., Gouder de Beauregard, V., Delaunoy, M., Jissendi, P., Cassart, M., & Désir, J. (2020). Middle interhemispheric variant of holoprosencephaly: First prenatal report of a ZIC2 missense mutation. *Clinical Case Reports*, 8(7), 1287–1292.  
<https://doi.org/10.1002/ccr3.2896>
- Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. In *Nature Reviews Molecular Cell Biology* (Vol. 19, Issue 10, pp. 621–637). Nature Publishing Group. <https://doi.org/10.1038/s41580-018-0028-8>
- Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., & Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology*, 42(2), 293–304. <https://doi.org/10.1038/s41587-023-01767-y>
- Haro, E., Petit, F., Pira, C. U., Spady, C. D., Lucas-Toca, S., Yorozya, L. I., Gray, A. L., Escande, F., Jourdain, A. S., Nguyen, A., Fellmann, F., Good, J. M., Francannet, C., Manouvrier-Hanu, S., Ros, M. A., & Oberg, K. C. (2021). Identification of limb-specific Lmx1b auto-regulatory modules with Nail-patella syndrome pathogenicity. *Nature Communications*, 12(1).  
<https://doi.org/10.1038/s41467-021-25844-5>
- Hébert, J. M., & Fishell, G. (2008a). The genetics of early telencephalon patterning: Some assembly required. In *Nature Reviews Neuroscience* (Vol. 9, Issue 9, pp. 678–685).  
<https://doi.org/10.1038/nrn2463>
- Hébert, J. M., & Fishell, G. (2008b). The genetics of early telencephalon patterning: Some assembly required. In *Nature Reviews Neuroscience* (Vol. 9, Issue 9, pp. 678–685).  
<https://doi.org/10.1038/nrn2463>
- Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., ... Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459(7243), 108–112. <https://doi.org/10.1038/nature07829>
- Hong, M., & Krauss, R. S. (2017). Ethanol itself is a holoprosencephaly-inducing teratogen. *PLoS ONE*, 12(4), 8–10. <https://doi.org/10.1371/journal.pone.0176440>

- Hossain, I., Priam, P., Reynoso, S. C., Sahni, S., Zhang, X. X., Côté, L., Doumat, J., Chik, C., Fu, T., Lessard, J. A., & Pastor, W. A. (2024). ZIC2 and ZIC3 promote SWI/SNF recruitment to safeguard progression towards human primed pluripotency. *Nature Communications*, 15(1), 8539. <https://doi.org/10.1038/s41467-024-52431-1>
- Houtmeyers, R., Souopgui, J., Tejpar, S., & Arkell, R. (2013). The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. *Cellular and Molecular Life Sciences*, 70(20), 3791–3811. <https://doi.org/10.1007/s00018-013-1285-5>
- Hsieh, T. H. S., Cattoglio, C., Slobodyanyuk, E., Hansen, A. S., Rando, O. J., Tjian, R., & Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Molecular Cell*, 78(3), 539–553.e8. <https://doi.org/10.1016/j.molcel.2020.03.002>
- Hua, P., Badat, M., Hanssen, L. L. P., Hentges, L. D., Crump, N., Downes, D. J., Jeziorska, D. M., Oudelaar, A. M., Schwessinger, R., Taylor, S., Milne, T. A., Hughes, J. R., Higgs, D. R., & Davies, J. O. J. (2021). Defining genome architecture at base-pair resolution. *Nature*, 595(7865), 125–129. <https://doi.org/10.1038/s41586-021-03639-4>
- Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R., & Bird, A. P. (2010). Orphan CpG Islands Identify numerous conserved promoters in the mammalian genome. *PLoS Genetics*, 6(9). <https://doi.org/10.1371/journal.pgen.1001134>
- Inoue, T., Ota, M., Mikoshiba, K., & Aruga, J. (2007). Zic2 and Zic3 synergistically control neurulation and segmentation of paraxial mesoderm in mouse embryo. *Developmental Biology*, 306(2), 669–684. <https://doi.org/10.1016/j.ydbio.2007.04.003>
- Iskusnykh, I. Y., Fattakhov, N., Li, Y., Bihannic, L., Kirchner, M. K., Steshina, E. Y., Northcott, P. A., & Chizhikov, V. V. (2023). Lmx1a is a master regulator of the cortical hem. *ELife*, 12. <https://doi.org/10.7554/ELIFE.84095>
- Jeong, Y., Leskow, F. C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., Dubourg, C., Li, X., Geng, X., Oliver, G., & Epstein, D. J. (2008). Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nature Genetics*, 40(11), 1348–1353. <https://doi.org/10.1038/ng.230>
- Kanton, S., Boyle, M. J., He, Z., Santel, M., Weigert, A., Sanchís-Calleja, F., Guijarro, P., Sidow, L., Fleck, J. S., Han, D., Qian, Z., Heide, M., Huttner, W. B., Khaitovich, P., Pääbo, S., Treutlein, B., & Camp, J. G. (2019). Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature*, 574(7778), 418–422. <https://doi.org/10.1038/s41586-019-1654-9>
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kim, S. H., & Chang, M. Y. (2023). Application of Human Brain Organoids—Opportunities and Challenges in Modeling Human Brain Development and Neurodevelopmental Diseases. In *International Journal of Molecular Sciences* (Vol. 24, Issue 15). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/ijms241512528>

- Kim, S., & Wysocka, J. (2023). Deciphering the multi-scale, quantitative cis-regulatory code. In *Molecular Cell* (Vol. 83, Issue 3, pp. 373–392). Cell Press. <https://doi.org/10.1016/j.molcel.2022.12.032>
- Lagutin, O. V., Zhu, C. C., Kobayashi, D., Topczewski, J., Shimamura, K., Puelles, L., Russell, H. R. C., McKinnon, P. J., Solnica-Krezel, L., & Oliver, G. (2003). Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes and Development*, 17(3), 368–379. <https://doi.org/10.1101/gad.1059403>
- Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D. J., Hicks, S. C., Robinson, M. D., Vallejos, C. A., Campbell, K. R., Beerenwinkel, N., Mahfouz, A., Pinello, L., Skums, P., Stamatakis, A., Attolini, C. S. O., Aparicio, S., Baaijens, J., Balvert, M., Barbanson, B. de, Cappuccio, A., ... Schönhuth, A. (2020). Eleven grand challenges in single-cell data science. In *Genome Biology* (Vol. 21, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13059-020-1926-6>
- Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. In *Cell* (Vol. 172, Issue 4, pp. 650–665). Cell Press. <https://doi.org/10.1016/j.cell.2018.01.029>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lavado, A., Lagutin, O. V., & Oliver, G. (2008). Six3 inactivation causes progressive caudalization and aberrant patterning of the mammalian diencephalon. *Development*, 135(3), 441–450. <https://doi.org/10.1242/dev.010082>
- Lê, S., Josse, J., Rennes, A., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. In *JSS Journal of Statistical Software* (Vol. 25). <http://www.jstatsoft.org/>
- Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. In *Cell* (Vol. 152, Issue 6, pp. 1237–1251). Elsevier B.V. <https://doi.org/10.1016/j.cell.2013.02.014>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, Y., Li, Z., Wang, C., Yang, M., He, Z., Wang, F., Zhang, Y., Li, R., Gong, Y., Wang, B., Fan, B., Wang, C., Chen, L., Li, H., Shi, P., Wang, N., Wei, Z., Wang, Y. L., Jin, L., ... Jiao, J. (2023). Spatiotemporal transcriptome atlas reveals the regional specification of the developing human brain. *Cell*, 186(26), 5892–5909.e22. <https://doi.org/10.1016/j.cell.2023.11.016>
- Lim, Y., & Golden, J. A. (2007). Patterning the developing diencephalon. In *Brain Research Reviews* (Vol. 53, Issue 1, pp. 17–26). <https://doi.org/10.1016/j.brainresrev.2006.06.004>
- Lo, H. F., Hong, M., & Krauss, R. S. (2021). Concepts in Multifactorial Etiology of Developmental Disorders: Gene-Gene and Gene-Environment Interactions in Holoprosencephaly. In *Frontiers in Cell and Developmental Biology* (Vol. 9). Frontiers Media S.A. <https://doi.org/10.3389/fcell.2021.795194>
- Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. In *Cell* (Vol. 167, Issue 5, pp. 1170–1187). Cell Press. <https://doi.org/10.1016/j.cell.2016.09.018>

- Long, H. K., Sims, D., Heger, A., Blackledge, N. P., Kutter, C., Wright, M. L., Grützner, F., Odom, D. T., Patient, R., Ponting, C. P., & Klose, R. J. (2013). Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *ELife*, 2013(2). <https://doi.org/10.7554/eLife.00348>
- Lopez-Rios, J. (2016). The many lives of SHH in limb development and evolution. *Seminars in Cell & Developmental Biology*, 49, 116–124. <https://doi.org/10.1016/j.semcdb.2015.12.018>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12). <https://doi.org/10.1186/s13059-014-0550-8>
- Lovely, C., Rampersad, M., Fernandes, Y., & Eberhart, J. (2017). Gene–environment interactions in development and disease. In *Wiley Interdisciplinary Reviews: Developmental Biology* (Vol. 6, Issue 1). John Wiley and Sons Inc. <https://doi.org/10.1002/wdev.247>
- Luo, Z., Gao, X., Lin, C., Smith, E. R., Marshall, S. A., Swanson, S. K., Florens, L., Washburn, M. P., & Shilatifard, A. (2015). Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Molecular Cell*, 57(4), 685–694. <https://doi.org/10.1016/j.molcel.2015.01.007>
- Machanick, P., & Bailey, T. L. (2011). MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics*, 27(12), 1696–1697. <https://doi.org/10.1093/bioinformatics/btr189>
- Mangale, V. S., Hirokawa, K. E., Satyaki, P. R. V., Gokulchandran, N., Chikbire, S., Subramanian, L., Shetty, A. S., Martynoga, B., Paul, J., Mai, M. V., Li, Y., Flanagan, L. A., Tole, S., & Monuki, E. S. (2008). Lhx2 Selector Activity Specifies Cortical Identity and Suppresses Hippocampal Organizer Fate. *Science*, 319(5861), 304–309. <https://doi.org/10.1126/science.1151695>
- Martinez-Ferre, A., & Martinez, S. (2012). Molecular regionalization of the diencephalon. In *Frontiers in Neuroscience* (Issue MAY). <https://doi.org/10.3389/fnins.2012.00073>
- Matsuda, K., Mikami, T., Oki, S., Iida, H., Andrabi, M., Boss, J. M., Yamaguchi, K., Shigenobu, S., & Kondoh, H. (2017). ChIP-seq analysis of genomic binding regions of five major transcription factors highlights a central role for ZIC2 in the mouse epiblast stem cell gene regulatory network. *Development (Cambridge)*, 144(11), 1948–1958. <https://doi.org/10.1242/dev.143479>
- Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kuttyavin, T., Stehling-Sun, S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., ... Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science (New York, N.Y.)*, 337(6099), 1190–1195. <https://doi.org/10.1126/science.1222794>
- McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., & Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, 28(5), 495–501. <https://doi.org/10.1038/nbt.1630>
- Merzdorf, C. S., & Sive, H. L. (2006). The zic1 gene is an activator of Wnt signaling. *International Journal of Developmental Biology*, 50(7), 611–617. <https://doi.org/10.1387/ijdb.052110cm>

- Meshorer, E., Yellajoshula, D., George, E., Scambler, P. J., Brown, D. T., & Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental Cell*, 10(1), 105–116. <https://doi.org/10.1016/j.devcel.2005.10.017>
- Michael Cohen, M., & Shiota, K. (2002). Teratogenesis of holoprosencephaly. *American Journal of Medical Genetics*, 109(1), 1–15. <https://doi.org/10.1002/ajmg.10258>
- Mizugishi, K., Hatayama, M., Tohmonda, T., Ogawa, M., Inoue, T., Mikoshiba, K., & Aruga, J. (2004). Myogenic repressor I-mfa interferes with the function of Zic family proteins. *Biochemical and Biophysical Research Communications*, 320(1), 233–240. <https://doi.org/10.1016/j.bbrc.2004.05.158>
- Morenilla-Palao, C., Teresa López-Cascales, M., López-Atalaya, J. P., Baeza, D., Calvo-Díaz, L., Barco, A., & Herrera, E. (2020). A Zic2-regulated switch in a noncanonical Wnt/ $\beta$ catenin pathway is essential for the formation of bilateral circuits. In *Sci. Adv* (Vol. 6). <https://www.science.org>
- Moreno-Bravo, J. A., Perez-Balaguer, A., Martinez, S., & Puelles, E. (2010). Dynamic expression patterns of Nkx6.1 and Nkx6.2 in the developing mes-diencephalic basal plate. *Developmental Dynamics*, 239(7), 2094–2101. <https://doi.org/10.1002/dvdy.22327>
- Mu, J., Zhou, Z., Sang, Q., & Wang, L. (2022). The physiological and pathological mechanisms of early embryonic development. In *Fundamental Research* (Vol. 2, Issue 6, pp. 859–872). KeAi Communications Co. <https://doi.org/10.1016/j.fmre.2022.08.011>
- Muzio, L., & Mallamaci, A. (2005). Foxg1 confines Cajal-Retzius neuronogenesis and hippocampal morphogenesis to the dorsomedial pallium. *Journal of Neuroscience*, 25(17), 4435–4441. <https://doi.org/10.1523/JNEUROSCI.4804-04.2005>
- Nagai, T., Aruga, J., Minowa, O., Sugimoto, T., Ohno, Y., Noda, T., & Mikoshiba, K. (2000). Zic2 regulates the kinetics of neurulation. *Proceedings of the National Academy of Sciences of the United States of America*, 97(4), 1618–1623. <http://www.ncbi.nlm.nih.gov/pubmed/10677508> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC26484>
- Nagai, T., Aruga, J., Takada, S., Gü, T., Spö, R., Schughart, K., & Mikoshiba, K. (1997). The Expression of the Mouse Zic1, Zic2, and Zic3 Gene Suggests an Essential Role for Zic Genes in Body Pattern Formation. In *DEVELOPMENTAL BIOLOGY* (Vol. 182).
- Naqvi, S., Kim, S., Hoskens, H., Matthews, H. S., Spritz, R. A., Klein, O. D., Hallgrímsson, B., Swigut, T., Claes, P., Pritchard, J. K., & Wysocka, J. (2023). Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. *Nature Genetics*, 55(5), 841–851. <https://doi.org/10.1038/s41588-023-01366-2>
- Nishimura, Y., & Kurosawa, K. (2022). Analysis of Gene-Environment Interactions Related to Developmental Disorders. In *Frontiers in Pharmacology* (Vol. 13). Frontiers Media S.A. <https://doi.org/10.3389/fphar.2022.863664>
- Oosterveen, T., Kurdija, S., Alekseenko, Z., Uhde, C. W., Bergsland, M., Sandberg, M., Andersson, E., Dias, J. M., Muhr, J., & Ericson, J. (2012). Mechanistic Differences in the Transcriptional Interpretation of Local and Long-Range Shh Morphogen Signaling. *Developmental Cell*, 23(5), 1006–1019. <https://doi.org/10.1016/j.devcel.2012.09.015>

- Özdemir, I., & Gambetta, M. C. (2019). The role of insulation in patterning gene expression. In *Genes* (Vol. 10, Issue 10). MDPI AG. <https://doi.org/10.3390/genes10100767>
- Pachano, T., Sánchez-Gaya, V., Ealo, T., Mariner-Faulí, M., Bleckwehl, T., Asenjo, H. G., Respuela, P., Cruz-Molina, S., Muñoz-San Martín, M., Haro, E., van IJcken, W. F. J., Landeira, D., & Rada-Iglesias, A. (2021). Orphan CpG islands amplify poised enhancer regulatory activity and determine target gene responsiveness. *Nature Genetics*, 53(7), 1036–1049. <https://doi.org/10.1038/s41588-021-00888-x>
- Parish, E. V., Mason, J. O., & Price, D. J. (2016). Expression of Barhl2 and its relationship with Pax6 expression in the forebrain of the mouse embryo. *BMC Neuroscience*, 17(1). <https://doi.org/10.1186/s12868-016-0311-6>
- Pourebahim, R., Houtmeyers, R., Ghogomu, S., Janssens, S., Thelie, A., Tran, H. T., Langenberg, T., Vleminckx, K., Bellefroid, E., Cassiman, J. J., & Tejpar, S. (2011). Transcription factor Zic2 inhibits Wnt/ $\beta$ -catenin protein signaling. *Journal of Biological Chemistry*, 286(43), 37732–37740. <https://doi.org/10.1074/jbc.M111.242826>
- Puelles, L. (2009). *Forebrain Development: Prosomere Model*.
- Puelles, L., & Martinez, S. (2013). Patterning of the Diencephalon. In *Patterning and Cell Type Specification in the Developing CNS and PNS* (pp. 151–172). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-397265-1.00048-4>
- Puelles, L., & Rubenstein, J. L. R. (2003). Forebrain gene expression domains and the evolving prosomeric model. *Trends in Neurosciences*, 26(9), 469–476. [https://doi.org/10.1016/S0166-2236\(03\)00234-0](https://doi.org/10.1016/S0166-2236(03)00234-0)
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Rada-Iglesias, A. (2014). Genetic variation within transcriptional regulatory elements and its implications for human disease. *Biological Chemistry*, 395(12), 1453–1460. <https://doi.org/10.1515/hsz-2014-0109>
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333), 279–285. <https://doi.org/10.1038/nature09692>
- Rallu, M., Machold, R., Gaiano, N., Corbin, J. G., McMahon, A. P., & Fishell, G. (2002). Dorsoroventral patterning is established in the telencephalon of mutants lacking both Gli3 and Hedgehog signaling. *Development*, 129(21), 4963–4974. <https://doi.org/10.1242/dev.129.21.4963>
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1), W160–W165. <https://doi.org/10.1093/NAR/GKW257>
- Rash, B. G., & Grove, E. A. (2007). Patterning the dorsal telencephalon: A role for sonic hedgehog? *Journal of Neuroscience*, 27(43), 11595–11603. <https://doi.org/10.1523/JNEUROSCI.3204-07.2007>

- Rossant, J., & Tam, P. P. L. (2022). Early human embryonic development: Blastocyst formation to gastrulation. In *Developmental Cell* (Vol. 57, Issue 2, pp. 152–165). Cell Press. <https://doi.org/10.1016/j.devcel.2021.12.022>
- Sagner, A., & Briscoe, J. (2017). Morphogen interpretation: concentration, time, competence, and signaling dynamics. *Wiley Interdisciplinary Reviews: Developmental Biology*, 6(4), 1–19. <https://doi.org/10.1002/wdev.271>
- Sapir, T., Sela-Donenfeld, D., Karlinski, M., & Reiner, O. (2022). Brain Organization and Human Diseases. In *Cells* (Vol. 11, Issue 10). MDPI. <https://doi.org/10.3390/cells11101642>
- Sato, T., Kikkawa, T., Saito, T., Itoi, K., & Osumi, N. (2017). Organizing activity of Fgf8 on the anterior telencephalon. *Development Growth and Differentiation*, 59(9), 701–712. <https://doi.org/10.1111/dgd.12411>
- Segert, J. A., Gisselbrecht, S. S., & Bulyk, M. L. (2021). Transcriptional Silencers: Driving Gene Expression with the Brakes On. In *Trends in Genetics* (Vol. 37, Issue 6, pp. 514–527). Elsevier Ltd. <https://doi.org/10.1016/j.tig.2021.02.002>
- Seidman, J. G., & Seidman, C. (2002). Transcription factor haploinsufficiency: when half a loaf is not enough. *Journal of Clinical Investigation*, 109(4), 451–455. <https://doi.org/10.1172/jci15043>
- Shimogori, T., Banuchi, V., Ng, H. Y., Strauss, J. B., & Grove, E. A. (2004). Embryonic signaling centers expressing BMP, WNT and FGF proteins interact to pattern the cerebral cortex. *Development*, 131(22), 5639–5647. <https://doi.org/10.1242/dev.01428>
- Sollier, E., Heilmann, J., Gerhauser, C., Scherer, M., Plass, C., & Lutsik, P. (2024). Figeno: multi-region genomic figures with long-read support. *Bioinformatics*, 40(6). <https://doi.org/10.1093/bioinformatics/btae354>
- Solomon, B. D., Lachawan, F., Mercier, S., Clegg, N. J., Delgado, M. R., Rosenbaum, K., Dubourg, C., David, V., Olney, A. H., Wehner, L. E., Hehr, U., Bale, S., Paulussen, A., Smeets, H. J., Hardisty, E., Tytki-Szymanska, A., Pronicka, E., Clemens, M., McPherson, E., ... Muenke, M. (2010). Mutations in ZIC2 in human holoprosencephaly: Description of a Novel ZIC2 specific phenotype and comprehensive analysis of 157 individuals. *Journal of Medical Genetics*, 47(8), 513–524. <https://doi.org/10.1136/jmg.2009.073049>
- Soluri, I. V., Zumerling, L. M., Parra, O. A. P., Clark, E. G., & Blythe, S. A. (2020). Zygotic pioneer factor activity of odd-paired/zic is necessary for late function of the drosophila segmentation network. *ELife*, 9. <https://doi.org/10.7554/eLife.53916>
- Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. In *Nature Reviews Genetics* (Vol. 13, Issue 9, pp. 613–626). <https://doi.org/10.1038/nrg3207>
- Spivakov, M. (2014). Spurious transcription factor binding: Non-functional or genetically redundant? *BioEssays*, 36(8), 798–806. <https://doi.org/10.1002/bies.201400036>
- Stark, R., & Brown, G. (n.d.). *DiffBind: Differential binding analysis of ChIP-Seq peak data*.
- Stevens, H. E., Smith, K. M., Rash, B. G., & Vaccarino, F. M. (2010). Neural stem cell regulation, fibroblast growth factors, and the developmental origins of neuropsychiatric disorders. In *Frontiers in Neuroscience* (Vol. 4, Issue SEP). <https://doi.org/10.3389/fnins.2010.00059>

- Stoykova, A., Treichel, D., Hallonet, M., & Gruss, P. (2000). *Pax6 Modulates the Dorsoventral Patterning of the Mammalian Telencephalon*.
- Suda, Y., Hossain, Z. M., Kobayashi, C., Hatano, O., Yoshida, M., Matsuo, I., & Aizawa, S. (2001). Emx2 directs the development of diencephalon in cooperation with Otx2. *Development*, 128(13), 2433–2450. <https://doi.org/10.1242/dev.128.13.2433>
- Suzuki, I. K., & Vanderhaeghen, P. (2015). Is this a brain which i see before me? Modeling human neural development with pluripotent stem cells. In *Development (Cambridge)* (Vol. 142, Issue 18, pp. 3138–3150). Company of Biologists Ltd. <https://doi.org/10.1242/dev.120568>
- Takagaki, Y., Yamagishi, H., & Matsuoka, R. (2012). Factors Involved in Signal Transduction During Vertebrate Myogenesis. In *International Review of Cell and Molecular Biology* (Vol. 296, pp. 187–272). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-394307-1.00004-7>
- Tanigawa, Y., Dyer, E. S., & Bejerano, G. (2022). WhichTF is functionally important in your open chromatin data? *PLoS Computational Biology*, 18(8). <https://doi.org/10.1371/journal.pcbi.1010378>
- Tchieu, J., Zimmer, B., Fattahi, F., Amin, S., Zeltner, N., Chen, S., & Studer, L. (2017). A Modular Platform for Differentiation of Human PSCs into All Major Ectodermal Lineages. *Cell Stem Cell*, 21(3), 399–410.e7. <https://doi.org/10.1016/j.stem.2017.08.015>
- Tekendo-Ngongang, C., Muenke, M., & Kruszka, P. (1993). *Holoprosencephaly Overview*.
- Tétreault, N., Champagne, M. P., & Bernier, G. (2009). The LIM homeobox transcription factor Lhx2 is required to specify the retina field and synergistically cooperates with Pax6 for Six6 trans-activation. *Developmental Biology*, 327(2), 541–550. <https://doi.org/10.1016/j.ydbio.2008.12.022>
- Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., ... Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature*, 489(7414), 75–82. <https://doi.org/10.1038/nature11232>
- Tian, E., Kimura, C., Takeda, N., Aizawa, S., & Matsuo, I. (2002). Otx2 is required to respond to signals from anterior neural ridge for forebrain specification. *Developmental Biology*, 242(2), 204–223. <https://doi.org/10.1006/dbio.2001.0531>
- Trapnell, C. (2015). Defining cell types and states with single-cell genomics. In *Genome Research* (Vol. 25, Issue 10, pp. 1491–1498). Cold Spring Harbor Laboratory Press. <https://doi.org/10.1101/gr.190595.115>
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5), 511–515. <https://doi.org/10.1038/nbt.1621>
- Ulloa, F., & Martí, E. (2010). Wnt won the war: Antagonistic role of Wnt over Shh controls dorso-ventral patterning of the vertebrate neural tube. In *Developmental Dynamics* (Vol. 239, Issue 1, pp. 69–76). <https://doi.org/10.1002/dvdy.22058>

- van der Lee, R., Correard, S., & Wasserman, W. W. (2020). Deregulated Regulators: Disease-Causing cis Variants in Transcription Factor Genes. In *Trends in Genetics* (Vol. 36, Issue 7, pp. 523–539). Elsevier Ltd. <https://doi.org/10.1016/j.tig.2020.04.006>
- Warr, N., Powles-Glover, N., Chappell, A., Robson, J., Norris, D., & Arkell, R. M. (2008). Zic2-associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation. *Human Molecular Genetics*, 17(19), 2986–2996. <https://doi.org/10.1093/hmg/ddn197>
- Wolpert, L. (1969). Positional Information and the Spatial Pattern of Cellular Differentiation. In *Theoret. Biol* (Vol. 25).
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation*, 2(3). <https://doi.org/10.1016/j.xinn.2021.100141>
- Xu, Y., Zhang, T., Zhou, Q., Hu, M., Qi, Y., Xue, Y., Nie, Y., Wang, L., Bao, Z., & Shi, W. (2023). A single-cell transcriptome atlas profiles early organogenesis in human embryos. *Nature Cell Biology*, 25(4), 604–615. <https://doi.org/10.1038/s41556-023-01108-w>
- Zeng, B., Liu, Z., Lu, Y., Zhong, S., Qin, S., Huang, L., Zeng, Y., Li, Z., Dong, H., Shi, Y., Yang, J., Dai, Y., Ma, Q., Sun, L., Bian, L., Han, D., Chen, Y., Qiu, X., Wang, W., ... Wang, X. (2023). The single-cell and spatial transcriptional landscape of human gastrulation and early brain development. *Cell Stem Cell*, 30(6), 851-866.e7. <https://doi.org/10.1016/j.stem.2023.04.016>
- Zhang, J., Hagopian-Donaldson, S., Serbedzija, G., Elsemore, J., Plehn-Dujowich, D., McMahon, A. P., Flavell, R. A., & Williams, T. (1996). Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2. *Nature*, 381(6579), 238–241. <https://doi.org/10.1038/381238a0>