2018-2019

Master's Degree in *Molecular Biology* and *Biomedicine*

Mechanistic characterization of ZIC2 function during brain patterning

Master's Thesis Project

Student: María Mariner Faulí

Institute: Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC)

Department: Signalling Department

Laboratory: Transcriptional Regulation in Development and Congenital Disease

Director: Dr. Álvaro Rada Iglesias

Table of contents

Abstract

Zinc finger of the cerebellum (ZIC) proteins constitute a family of transcription factors (TFs) with crucial roles during embryogenesis, particularly during neural development. Defects on the genes encoding these TFs cause a broad range of developmental disorders. In particular, *Zic2* defects lead to holoprosencephaly, a congenital brain malformation resulting from the defective cleavage of cerebral hemispheres manifested with variable expressivity and incomplete penetrance. However, the target genes and mechanism of action of ZIC2 during brain development are largely unknown. Consequently, the molecular etiology of ZIC2-associated holoprosencephaly remains poorly characterized. To elucidate the molecular mechanisms by which ZIC2 contributes to proper brain development, I first analyzed how the loss of ZIC2 function affects the differentiation of embryonic stem cells into anterior neural progenitor cells (AntNPCs). Notably, I found that the knockout of *Zic2* led to a drastic downregulation of dorsal brain genes, including major roof plate markers such as *Lmx1a* and *Lmx1b*. Next, one major objective in this project is to determine if ZIC2 directly activates these dorsal genes during AntNPC differentiation or if, alternatively, it represses ventral regulators which themselves antagonize brain dorsal identity. To achieve this, it is necessary to generate ZIC2 binding profiles genome-wide in AntNPC. Due to the lack of a specific antibody against ZIC2, I used CRISPR-Cas9 technology to generate a mouse embryonic stem cell line (mESC) in which the endogenous *Zic2* was tagged with a C-terminal Flag-HA epitope. After demonstrating that *Zic2* is expressed in this cell line at the same levels as in WT cells both at mRNA and protein level, I also showed that this cell line can be used to identify ZIC2 binding sites by chromatin immunoprecipitation (ChIP). This *Zic2*-Flag-HA mESC line will now allow us to perform chromatin immunoprecipitation and sequencing (ChIP-seq) and immunoprecipitation coupled to mass spectrometry (IP-MS) experiments to elucidate ZIC2 genomic binding sites and its possible interacting partners, which should provide major insights into the regulatory networks and mechanisms whereby ZIC2 contributes to brain development and human holoprosencephaly.

Acknowledgements

After all the work that this project recapitulates I could not leave unmentioned all the people that has been involved in making this Master's Thesis an awesome experience both at the scientific and personal level:

I would like to thank Dr. Álvaro Rada for accepting me in his team and for his close supervision and guidance. I must admit that he reaches all expectations as a PI and I still do not believe how lucky I am of working and learning each day in this lab.

Thanks to Patricia, "the bench master", for every advice and help she has given me since I arrived and for making easier each day with our funny conversations. You need her, she is there.

Thanks to Sara, my first supervisor in Cologne. She started this project and I will work hard to meet the expectations so as she can be proud of her little Padawan. I miss you, a lot. Thanks for teaching me and taking care of me at the very beginning. And for the memes, best of it.

Thanks to Sarah, my French best friend, for all the time spent in and outside the IBBTEC. For the mutual moral support and the stupid jokes late in the lab.

Thanks to Dra. Marian Ros for her constant advice on developmental biology. Each video she has sent me and each explanation she has given me have made this complex world a bit easier.

Thanks to all the unique and crazy members of Marian Ros' lab. Love you, vaquitas.

Thanks to Piero's lab girls for the help with the Western Blot and for their continuous good vibes.

Thanks to all the master colleagues, specially Aurora, Ale and Maria, the discovery of the year.

Thanks to my parents for all the effort and support during all my education. But mainly for raising a human who loves nature, because that is the main reason why I am here.

Thanks to Dundee crocodile, Manel. You are the best brother anyone could ask for.

And thanks to Víctor, my favourite person in the world. Everything is better and funnier by your side. Impossible to put in words how grateful I am to spent each day with you.

Abbreviations

ZIC: Zinc Finger of the Cerebellum

TF: Transcription Factor

HPE: Holoprosencephaly

AntNPCs: Anterior Neural Progenitor cells

mESCs: Mouse embryonic stem cells

ChIP: Chromatin Immunoprecipitation

IP: Immunoprecipitation

MS: Mass Spectrometry

MIHV: Middle Interhemispheric variant

RT-qPCR: Real Time quantitative polymerase chain reaction

WB: Western Blot

PE: Poised Enhancer

LIF: Leukemia Inhibiting Factor

PBS: Phosphate Buffered Saline

HRP: Horseradish peroxidase

gDNA: Genomic DNA

IF: Immunofluorescence

IP-MS: Immunoprecipitation coupled to mass spectrometry

COS: Cells being **C**V-1 (simian) in **O**rigin, and carrying the **S**V40 genetic material

dpc: Days post coitum

SDS-PAGE: Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

PVDF: Polyvinylidene fluoride

BSA: Bovine serum albumin

1. Introduction

1.1. ZIC family of transcription factors

Zinc finger of the cerebellum genes (ZIC) encode a family of transcription factors (TFs) with a critical role during embryogenesis [1]. They are orthologues of the *Drosophila melanogaster* odd-paired gene, which due to an evolutionary gene copy expansion resulted in the five ZIC homologs that are now present in both mouse and human (*i.e. Zic1* to *Zic5*). All ZIC proteins share a zinc finger domain consisting of five highly conserved Cys2His2-type zinc fingers [2,3] that can bind both DNA and other proteins [4]. ZIC proteins act at different stages of early (neurulation, neuroectodermal differentiation or neural crest induction) [1,4] and late neural development (axon guidance, refinement of axon terminals) [5,6]. The identification of ZIC mutations in different congenital syndromes highlights the relevance of these genes in human development [4].

1.2. ZIC2-associated human holoprosencephaly (HPE)

In this project, we focus on one member of this TF family, ZIC2. Loss of function mutations within the human *ZIC2* gene are associated with holoprosencephaly (HPE). HPE is a developmental brain malformation resulting from incomplete cleavage of the prosencephalon and affecting the correct separation of the two brain hemispheres and facial structures [7]. Together with *SHH*, *SIX3* and *TGIF*, *ZIC2* is one of the four marker genes screened for the molecular diagnosis of new sporadic or familial HPE cases [7–9]. However, to date, fourteen genes have been already associated with this malformation in humans (*SHH*, *ZIC2*, *TGIF*, *SIX3*, *CDON*, *DISP1*, *DLL*, *FGF8*, *FGFR1*, *FOXH1*, *GAS1*, *PTCH1*, *NODAL*, *TDGF1*) [10]. Moreover, HPE displays reduced penetrance and variable expressivity: not all individuals with loss of function mutations in the previous genes display HPE and the severity of the brain abnormalities differ between patients [7]. HPE is considered as a rare disease, which, nevertheless, is the most common brain structural defect in humans, with an incidence of 1/250 conceptuses and 1/10.000-16.000 live births [7–9].

HPE can be broadly classified in two major types attending to the degree of brain separation and to the location of the defect with respect to the dorso-ventral axis: classical HPE and middle interhemispheric variant (MIHV).

In classical HPE the lack of separation is most severe ventrally, and three subgroups are established following a severity criterion: (i) alobar (monoventricle with no hemisphere separation), (ii) semilobar (partial hemispheric separation), and (iii) lobar (hemispheric and lateral vesicle separation is preserved but absent in the rostral and ventral frontal lobes). These

phenotypes also include craniofacial defects as microcephaly, hypotelorism and/or cleft lip and palate, whose severity normally correlates with the degree of hemispheres separation. In addition, midline defects, such as undivided thalami, absent corpora callosa and absent or hypoplasic olfactory and optic bulbs, are also frequent [10].

MIHV, in contrast to classical HPE, shows a proper separation of the ventral forebrain but presents a defect in the division of the posterior frontal and parietal regions of the cerebral hemispheres along the dorsal midline [11]. Both classic and MIHV HPE present an impairment in the cleavage of cerebral hemispheres, however, MIHV is milder and less frequent than classic HPE [12] and some authors propose that they may have a distinct embryological origin [13].In one study that supports this hypothesis, mouse embryos defected for the BMP receptor genes *Bmpr1b* and *Bmpr1a* (*Bmpr1b*$^{-/-}$/*Bmpr1a*$^{-/-}$ mice) displayed a loss of all dorsal midline cell types without affecting the specification of cortical and ventral precursors, in opposition to *Shh*$^{-/-}$ mutants, in which ventral patterning is disrupted, but the dorsal midline initially forms. However, assigning a role to ZIC2 in the process of hemisphere separation still remains a challenge, as *Zic2* is the only gene associated to both classical and MIHV HPE [10].

In mouse model, spontaneous null mutations (*Zic2 Kumba allele*) demonstrated that the complete absence of ZIC2 (*Zic2*$^{Ku/Ku}$) causes a mid-gastrulation failure that leads to classical HPE with variable severity[14]. In addition, a hypomorphic *Zic2* mouse line, with approximately 20% of the normal *Zic2* expression levels, develops normally though gastrulation, but displays dorsal forebrain malformations at later stages of development including MIHV HPE, microcephaly, exencephaly and spina bifida at different degrees of severity [15]. These data on mice models confirm that ZIC2 associated HPE manifests a defective forebrain dorso-ventral patterning. Interestingly, the defects associated to an absent (*Zic2*$^{Ku/Ku}$) versus a low level of *Zic2* expression (hypomorphic line, 20% of activity) suggest that *Zic2* dosage sensitivity might change during embryogenesis, with dorsal brain patterning being particularly sensitive to reduced ZIC2 levels.

As previously mentioned, HPE, as many other congenital syndromes, is characterized by its variable expressivity and incomplete penetrance [16]. These two features strongly suggest that HPE might not be simply caused by loss of function mutations in the relevant genes (including *ZIC2*), but interactions with additional genetic and environmental factors might also contribute to the etiology of this disease.

1.3. Gene-environmental interactions

Studies in mice show that homozygous *Zic2* hypomorphic mutants frequently display forebrain abnormalities and die at perinatal stages, but heterozygous mice do not present major abnormalities [15]. In comparison, the reported human cases appear due to *Zic2* heterozygous loss-of-function mutations or deletions. The fact that heterozygous mice do not present major brain anomalies, but human reported cases are caused by heterozygous mutations could suggest different gene dosage sensitivity between mice and humans. However, another possibility is that interactions between *ZIC2* mutations and environmental risk factors (i.e. geneXenvironmental interactions (GxE)) cause these differences between species. Namely, laboratory mice develop under highly controlled and uniform conditions, whilst environmental context during human gestation is highly variable. Exposure to ethanol and retinoic acid during development are the most well characterized risk factors for HPE [17]. Both of these teratogenic factors are believed to cause HPE by disrupting dorso-ventral brain patterning, although the precise mechanism of action has not been fully elucidated. Interestingly, exposure to these teratogens does not cause HPE with full penetrance, further supporting that additional interaction of these compounds with genetic risk factors cause HPE [8,16]

1.4. *Zic2* expression in mouse

To date, most of the information obtained about ZIC2 function is based on the study of this TF during mouse embryogenesis. *Zic2* expression domains during mouse development have been characterized by *in situ* hybridization and immunofluorescence techniques (IF). *Zic2* expression initiates at very early stages of development, being detectable in the early embryo stage of fertilized zygotes and in pre-implantation blastocysts [18]. Prior to gastrulation, it is also expressed in the ectoderm of both the extraembryonic and embryonic components of the egg cylinder at 5.5 days post coitum (dpc). As gastrulation progresses (aprox. 7.0 dpc), *Zic2* transcripts are detected not only in the ectoderm, but also in the primitive streak and adjacent cells forming the node and in the emerging mesoderm in the distal two thirds of the embryonic region. By the neural plate stage, *Zic2* is only expressed in the anterior half of the embryo and *Zic2* signal is predominantly detected in the neural plate region. As neurulation proceeds, *Zic2* expression becomes restricted to the most dorsal region in the cranial neurectoderm and subsequent neural tube [19].

In order to better understand the results of this project, it is important to explain the main characteristics of neural tube formation. Towards the end of gastrulation, the neural plate begins to fold and forms the neural tube. The cells of the most anterior part of the neural tube will give rise to the brain and the posterior ones will give rise to the spinal cord, together forming the central nervous system. In addition to distinct anterior-posterior identities, the neural tube also undergoes dorso-ventral patterning in response to specific signals (Fig.3.A). Ventral identities are induced by Sonic Hedgehog (SHH) signaling emanating from the notochord and the neural tube floor

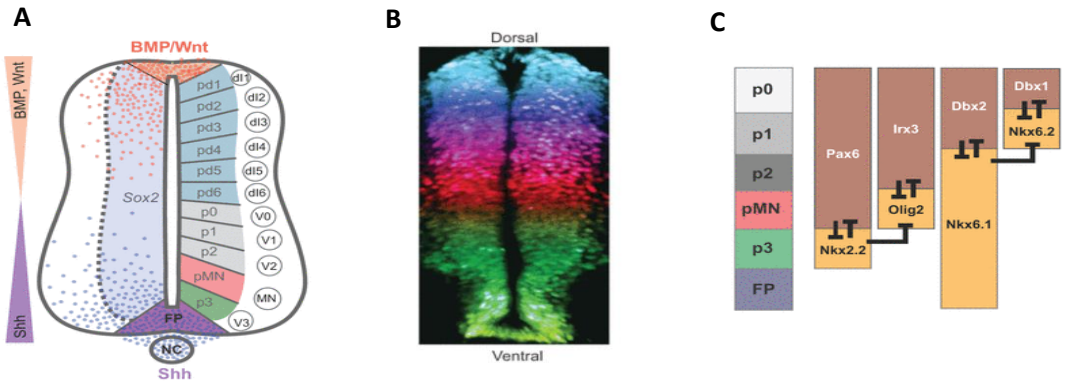

Figure 3. Dorsal-ventral patterning of the vertebrate neural tube. **A**. Progenitor domain (pd1-pd6, p0-p3) identity is based on the combinatorial expression of a set of TFs and this combinatorial code is necessary and sufficient to specify the neuronal subtypes (V0-V3, MN, dI1-dI6) that each domain generates. The pattern of gene expression is established in a progressive manner in response to opposing gradients of secreted factors: Shh emanating from the ventral pole (NC, notochord); Wnt and BMP signaling dorsally. **B.** Transversal section of a mouse neural tube. Combination of different immunofluorescence images to illustrate the expression of the TFs marking each progenitor domain. **C.** Illustration of the mutually repressive interactions established among TF of adjacent ventral progenitor domains. Modified from *Briscoe and Small 2015*.

plate, whereas dorsal identities are induced by WNTs and BMPs signals coming from the roof plate. Ventral and dorsal signals antagonize each other's action, creating an opposing gradient of concentrations that will be interpreted by the cells of the neural tube. Depending on its position, each cell will interpret different inputs of information. This will allow the specification of different neuronal progenitor domains along the transversal section of the neural tube, each of them characterized by the expression of a unique set of transcription factors (Fig.3.B). Subsequently, the boundaries between these progenitor domains become more sharped due to the mutually repressing interactions between the TFs expressed in adjacent domains (Fig.3.C). Ultimately, this will result in the establishment of different transcriptional programs in each progenitor domain.

1.5. Embryonic Stem Cells (ESC) as a tractable *in vitro* model

Mouse studies have demonstrated the important role of ZIC2 during brain development. However, there is still a limited understanding of the gene regulatory networks controlled by ZIC2 as well as of the regulatory mechanism by which this TF controls gene expression. Consequently, the etiological mechanisms by which ZIC2 mutations cause HPE are still unknown, limiting the

implementation of novel diagnostic or therapeutic strategies. Our project aims to shed some light into these questions. To do so, it is necessary to apply different genomic approaches, such as chromatin immunoprecipitation and sequencing (ChIP-seq) and RNA-seq, to interrogate ZIC2 function at different stages of neural differentiation. For an optimal performance of these techniques, high amounts of cells are required, and currently it is technically very challenging to obtain the specific cell types (i.e. neural progenitors) in sufficient amounts *in vivo*. Moreover, there are obvious ethical restrictions on performing these methodologies on human embryos. To solve this issue, mouse embryonic stem cells (mESCs) represent a highly tractable and powerful tool as an *in vitro* model to study early differentiation processes and to model congenital diseases. Our team has previously optimized a 5-day differentiation protocol of mESCs into anterior neural progenitor cells (AntNPCs) that models the earlier stages of mouse brain development [20].

1.6. ZIC2 as a Transcriptional Regulator

Applying different genomic and genetic engineering approaches to the mESC differentiation model described above, our team previously uncovered poised enhancers (PE) as a group of highly conserved regulatory elements that are in a "pre-marked" but inactive state in ESC. Importantly, PE contribute to the establishment of a genetic program essential for early brain development as they become active in neural progenitors [20]. However, it is currently unknown how PE become activated during anterior neural induction. Motif analysis in PE sequences revealed a significant enrichment of ZIC binding sites. In addition, upon differentiation of mESC into anterior neural progenitors (AntNPC), we observed that although *Zic2* was already expressed in mESC, its expression increased upon AntNPC differentiation. Moreover, analysis of public ChIP-seq data revealed that a significant fraction of PE is already bound by ZIC2 in mESC. All together, these observations made us hypothesize that ZIC2 could be involved in the activation of a subset of PE and, thus, in the proper establishment of anterior neural identity.

Research made by Luo et al. [21] shows that after silencing ZIC2 in mESC by shRNA there are important gene expression changes. Moreover, upon differentiation of these cells into neural progenitors, neural linage commitment was abrogated and massive cell death was observed, suggesting that ZIC2 could be important for cellular pluripotency. In addition, they used co-immunoprecipitation techniques to prove that ZIC2 could bind MBD3, one of the main members of the nucleosome remodeling complex NuRD, a unique chromatin-remodeling complex with both chromatin opening and closing activities that is mostly considered as a co-repressor [21,22]. However, using CRISPR/Cas9 technology, our team recently generated *Zic2*$^{-/-}$ mESC that did not display any major gene expression or phenotypic defects. Moreover, we were able to differentiate *Zic2*$^{-/-}$ mESC into AntNPCs without any obvious cell death, or morphological defects. This

indicates that the defects observed by Luo et al. could be due to off-target and/or acute effects of the shRNA technique used in their study. In addition, our team has also generated RNA-seq data during the differentiation of either WT or *Zic2*$^{-/-}$ mESC into AntNPC. When considering undifferentiated ESC, the RNA-seq data revealed rather minor gene expression changes between WT and *Zic2*$^{-/-}$ cells. In contrast, gene expression differences were very pronounced in AntNPC, with a clear downregulation of dorsal genes and an upregulation of ventral genes in *Zic2*$^{-/-}$ in AntNPCs.

On the other hand, previous work in our laboratory showed that, in our hands, the commercial ZIC2 antibody used in Luo et al. to generate ChIP-seq data in mESC was not completely specific for ZIC2 when used for either ChIP or IF. Therefore, and due to the lack of any other commercial ChIP-Grade ZIC2 antibody, a major aim of this project was to use CRISPR-Cas9 technology to generate a mESC line stably expressing the endogenous ZIC2 tagged with a C-terminal Flag-HA epitope. In principle, this mESC line should allow us to perform ChIP-seq, IP-MS (immunoprecipitation coupled to mass-spectrometry) experiments using antibodies against the inserted epitopes and avoiding unspecificity issues frequently associated with the use of commercial antibodies. These experiments should help us to uncover ZIC2 binding regions and its possible protein interacting partners upon differentiation of mESC into AntNPCs. Consequently, these experiments could provide major insights into the regulatory mechanisms whereby ZIC2 contributes to the establishment of an anterior neural gene expression program.

Currently, all the available data point out towards ZIC2 representing a multi-functional transcriptional regulator at different stages of embryonic development. Nevertheless, there is still a lack of global data that relate ZIC2 binding sites with transcriptional or epigenetic changes during neural development. This information combined with the one obtained from studying the effects of *Zic2* knock out during embryogenesis could help us to understand ZIC2 function in the context of normal brain development as well as in the etiology of HPE. Following this objective, in this project we further analyzed the expression profiles of *Zic2*$^{-/-}$ mESC during AntNPC differentiation. Consistent both with previous studies in mouse embryos [13] and with the RNA-seq data previously obtained by our team, I confirmed that the loss of ZIC2 during the differentiation of ESC into AntNPC causes an upregulation of ventral genes and downregulation of dorsal ones. Among these gene expression changes, I observed very severe defects in the expression of the major roof plate regulators *Lmx1a* and *Lmx1b*. To ask whether ZIC2 is directly involved in the regulation of some of these genes, and considering the absence of a specific ChIP-grade antibody for ZIC2, I successfully generated a *Zic2*-Flag-HA tagged mESC line as a research tool for future ChIP-Seq and IP-MS experiments. This mESC line should enable us to interrogate ZIC2 genomic

binding regions and uncover its possible protein interacting partners. Finally, using this newly generated mESC line, I already optimized conditions for future ChIP experiments and confirmed that this cell line can be used to map ZIC2 binding sites.

2. Objectives

1. Validation of the gene expression differences previously identified in *Zic2*$^{-/-}$ vs WT AntNPCs.
2. Generation of a *Zic2*-Flag-HA tagged mESC line to globally map ZIC2 genomic binding regions and to identify its possible interacting partners.
3. Optimization of ChIP conditions on the newly generated *Zic2-Flag-HA* mESC line.

3. Materials and methods

2.1. Mammalian cell culture procedures

Cell culture protocols were performed under sterility conditions guaranteed by laminal flow cell culture hoods, sterile solutions and media supplemented with antimitotics and antibiotics. Cells were kept in an incubator at 37° in a humid 5% CO2 atmosphere.

2.1.1. Mammalian cell lines

For this project, two types of mESC lines were used. In addition, one Flag-HA Zic2-tagged cell line was generated as part of the project:

Table 1. Cell lines used in the current project.

Cell line	Reference
WT (E14) mESC	Wysocka Laboratory
Zic2-/- mESC	Rada-Iglesias Laboratory (already available)
Zic2-Flag-HA mESC (FH6 and FH12)	Rada-Iglesias Laboratory (generated during this project)
COS cells	Rada-Iglesias Laboratory (already available)

2.1.2. Culture of mESC in serum plus leukemia inhibition factor (LIF) conditions

All cell lines used in this study were grown on 1% gelatin-coated plates using knock-out DMEM (KO-DMEM, Life Technologies) supplemented with 15% FBS (Heat Inactivated FBS, Gibco,

10500) and LIF (see 2.1.3. Production of LIF-1ca from COS cells) in order to promote a primed pluripotency state.

2.1.3. Production of LIF-1ca from COS cells

COS cells were grown in complete mESC growing media lacking LIF and were expanded into 15 cm plates. When they reached confluence, their growing media was collected over 2 nights and stored at 4°C. Then, cells were spun down at 1000 rpm for 5 min from the media previously collected and the supernatant was filtered through a 0.2µm filter. All this media rich in LIF was aliquoted and tested prior to use, to adjust the optimal working concentration.

2.1.4. Differentiation into Anterior Neural Progenitors (AntNPCs)

Cells were initially plated at a density of 18.000 cells/cm^2 on overnight 1% gelatin-coated plates in N2B27 medium supplemented with 1mg/ml Bovine Serum Albumin (BSA) and 10ng/ml β-FGF (Human recombinant β-FGF, Gibco, 13256029) without serum nor LIF (D0), following a previously stablished protocol with slight modifications [20]. The two next consecutive days, BSA dose was reduced to 40µg/ml, β-FGF was kept at 10ng/ml and at D2 media was also supplemented with 5µM Xav939. Then, cells were grown for three more days without β-FGF but supplemented again with 5µM Xav939 in order to improve the homogeneity of the differentiation by enriching ectodermal fates [23].

Table 2. Supplements added each day (D1-D5) of the differentiation of mESCs into AntNPCs. S: Supplement, C: concentration.

D0		D1		D2		D3		D4		D5	
S	**C**	**S**	**C**	**S**	**C**	**S**	**C**	**S**	**C**	**S**	**C**
BSA	1mg/ml	BSA	40µg/ml	BSA	40µg/ml	BSA	40µg/ml	BSA	40µg/ml	Final day. RNA extraction.	
β-FGF	10ng/ml	β-FGF	10ng/ml	β-FGF	10ng/ml	Xav939	5µM	Xav939	5µM		
				Xav939	5µM						

2.2. Molecular biology methods.

2.2.1. Genomic DNA isolation

DNA of mESCs was isolated using "Quick genomic DNA extraction protocol" Lysis Buffer recipe [24]. Approximately, 0.5 ml of Lysis Buffer were added per 10^4 cells in a microcentrifuge 1.5ml Eppendorf tube containing only the cellular pellet. Then, the pellet was properly resuspended by vortexing for 15 sec and 1µl of Proteinase K (20µg/µl) for every 25µl of Lysis Buffer was added. This step was followed by an incubation of 6 min at 65°C to allow the digestion of proteins present in the sample. Then, another vortexing step was performed again for 15 sec

followed by Proteinase K inactivation at 98°C for 2 min. After that, genomic DNA is ready to use or storage.

2.2.3. RNA isolation

Total RNA was isolated using SPEEDTOOLS Total RNA Extraction Kit (Biotools B&M Labs S.A.) following manufacturer's instructions.

2.2.4. cDNA synthesis

To perform expression analyses, RNA was reversely transcribed into cDNA using ProtoScript II First Strand cDNA Synthesis Kit. Exclusive transcription of mRNAs was warranted by the use of oligo-dT Primer binding to the mRNA specific poly-A tail and by the design of primers whose PCR product covered two exons.

2.3. Immunological methods

2.3.1. ChIP

To perform ChIP experiments, a previously described protocol[25] was followed with slight modifications. Initially, sonication conditions were optimized (see 3. Results). Briefly, $5x10^7$ cells for DNA Polymerase II ChIP, $1x10^7$ cells for H3K4me3 ChIP and $11x10^7$ cells for ZIC2 ChIP were crosslinked with 1% formaldehyde for 10 minutes at RT and then quenched with 0,125M glycine for another 10 min. Then, cells were rinsed with phosphate buffered saline (PBS, Sigma, RNBG8633) and resuspended sequentially in three different lysis buffers (Table 3) to isolate chromatin. Chromatin was then sonicated for 15 cycles (cycle = 30 sec on/ 30 sec off; Amplitude 80%) using Ultrasonic processor (Labsonic). After sonication, the suspension was centrifuged 10 min at 16000g and 4°C. Afterwards, the chromatin from the supernatant was divided in different aliquots that were incubated overnight at 4°C with 7 ug of antibody for H3K4me3, 10 ug of antibody for Pol2, Flag and HA ChIP. A 10% volume of each aliquot for ChIP was incubated without the antibody as a representation of the total input control for the ChIP reactions.

Table 3. Lysis buffers used for ChIP.

Lysis Buffer	Composition
1	50mM HEPES, 140 mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X-100
2	10mM Tris-HCl ph 8.0, 200 mM NaCl, 1mM EDTA, 0.5 mM EGTA
3	10mM Tris-HCl, 100mM NaCl, 1mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine

Next day, 100µl of protein G magnetic beads were added to the Pol2 and ZIC2 ChIP reactions and 75µl to the histone ChIP. After four-hour incubation at 4°C, magnetic beads were washed

with RIPA wash buffer (50mM HEPES, 500mM LiCl, 1mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate) and chromatin eluted, followed by reversal of the crosslinking and DNA purification. Briefly, for DNA purification, two extractions were performed, one with phenol/chlorophorm/isoamyl alcohol (25:24:1), followed by a second chlorophorm extraction. The aqueous phase was isolated and DNA was precipitated during 30 min at -80°C by adding 1/10 of the volume of sodium acetate, 1µl glycogen (as internal carrier) and 3 volumes of 100% ethanol. Afterwards, samples were centrifuged during 30 min at 4°C with DNA representing pellet. DNA was eluted in water. All antibodies used have been previously reported as ChIP-grade (Table 4). ChIP samples were analyzed by q-PCR using the primers shown in Table 11.

Table 4. Antibodies used for ChIP.

Antibody	Company	Reference
DNA Polymerase II	Millipore	8WG16
H3K4me3	Active motif	39159
Flag epitope	Sigma	F1804-1MG
HA epitope	Abcam	Ab9110

2.3.2. Western Blot

To confirm that our cell lines expressed a fusion protein composed by ZIC2 and the Flag-HA epitope, Western Blot was performed. Proteins from WT, FH6 and FH12 mESCs were extracted using RIPA buffer (50mM Tris HCl pH 7.5, 150mM NaCl, 1mM EDTA, 1% (v/v) NP40, 0.1% SDS, 0.5% sodium deoxycholate) supplemented with protease inhibitor cocktail. After 20 minutes of incubation on ice, protein extracts were recovered by centrifugation (20 minutes at 14000g). For Western Blot, 150µg of protein were mixed with Laemmli buffer, heated to 95°C and then separated in a 15% SDS-PAGE gel in running buffer (25mM Tris base, 250 mM glycine, 0.1% SDS). Proteins were transferred to a PVDF (polyvinylidene difluoride) membrane using transfer buffer for 1 h at 100V. The membranes were blocked for 1 h in 4% (w/v) BSA powder and incubated with primary antibody overnight at 4°C. After 3-5 washes with TBST, the membranes were incubated with the secondary antibody for 1 h at room temperature. Horseraddish peroxidase (HRP) coupled anti-IgG antibody was detected using a chemiluminiscent substrate. Antibodies used are listed in Tables 5 and 6.

Table 5. Primary antibodies used for Western Blot.

Antigen	Host	Dilution	Company	Reference
HA	Rabbit	1:5000	Abcam	Ab9110
ZIC2	Rabbit	1:1000	Abcam	ab150404
H3K27ac	Rabbit	1:10000	Active motif	39133

Table 6. Secondary antibody used for Western Blot.

Species	Conjugated	Dilution	Company	Reference
Anti-rabbit	HRP	1:5000	Invitrogen	656120

2.4. Genetic engineering methods

2.4.1. Generation of a ZIC2 Flag-HA tagged cell line using CRISPR-Cas9

2.4.1.1. Design of guide RNA (sgRNA)

In order to generate a mESC line with the endogenous ZIC2 gene tagged with the Flag-HA epitope in the C-terminal end, a sgRNA was designed to allow the cut just before *Zic2* transcriptional stop codon. The selection of the optimal sgRNA was performed following Benchling CRISPR toolbox instructions (https://www.benchling.com/crispr/). This sgRNA and its complementary oligo (Table 7) were synthesized, both carrying at the end a 4 nucleotide sequence complementary to BbsI restriction sites, as it was the enzyme used to digest the vector pX330A_hCas9_long_chimeric_gRNA_G2P (Fig.1). The two synthesized oligos were annealed by incubation at 95° for 5 min and subsequent cooling to 25° at a cooling rate of 5°/min, following the next combination of reagents:

Mix Ligation reaction. *1X (10µl):*

1µl oligonucleotide A

1µl oligonucleotide A'

1µl 10X T4 Ligation Buffer (New England Biolabs)

7µl ddH2O

Table 7. gRNA oligos designed.

gRNA oligo F (5'-3')	gRNA oligo R (5'-3')
caccCAATGAATGGTACGTGTGAG	aaacCTCACACGTACCATTCATTG

2.4.1.2. Generation of guide RNA (sgRNA) Cas9 vector

The CRISPR-Cas9 expression vector pX330A_hCas9_long_chimeric_gRNA_G2P (Fig. 1) was digested with BbsI restriction enzyme and column purified with SPEEDTOOLS PCR CLEANUP KIT (Biotools, 21.202)

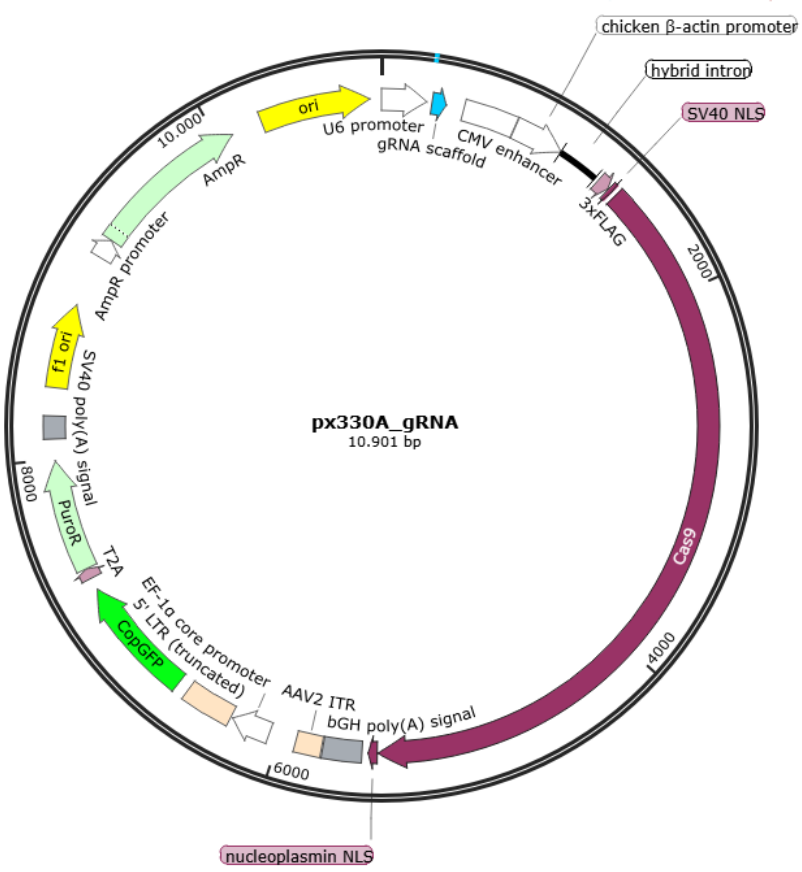

Figure 1. Scheme of pX330A_hCas9_long_chimeric_gRNA_G2P vector used for mESC transfection.

Mix Digestion reaction. *1X*

1µg vector

1µl BbsI

2µl Buffer

Xµl ddH2O

A dilution 1:200 of the pair of annealed oligos and 50 ng of the digested vector were ligated overnight at 16° using T4 ligase (New England Biolabs). Then, chemical competent *E. coli* were transformed by heat shock. A 30µl aliquot of bacteria was thaw on ice and mixed with 2.5µl of

the ligation reaction. After inverting the tube three times, bacteria solution was incubated 1 minute on ice, followed by heat shock at 42° for 25 secs and an extra 1 min on ice. Subsequently, the solution was transferred to 1ml LB medium and incubated 1 hour at 37° with shaking. 200µl of this liquid culture were plated in LB ampicillin plates in order to obtain bacteria colonies which were analyzed by PCR to determine if they carried the plasmid of interest. See 2.5.3 Colony PCR.

2.4.1.4. Design of a repair template including Flag-HA

To insert the Flag-HA epitope at the site where Cas9 endonuclease introduced the cut, a repair template was designed to induce homology directed repair. This sequence was composed by two 75 bp homology arms, complementary to both sides of the cut at the gDNA flanking the sequence of the Flag-HA epitope followed by a transcriptional stop codon (Homology arm left: 5'-GGGGCGGGCGGCGGGGGCGGCGGCAGCTCTGGCGGGGGCAGCGGGACAACCGGAGGCCATAGCGGCCTCTCCTCCAACTTCAATGAATGGTACGTG/Epitope:GACTACAAAGACGACGACGACAAATACCCATACGACGTGCCAGACTACGCC/ Stop:TGA / Homology arm right: GGGGCCAGGCCTTTCTCCCATTCCCTGTTCCCTTATCCACCGTCGCCCTCCCAAAACCCATCGAGGGCACCTTAGGATCGTCTTATTAAAATTATG-3'). The single stranded repair sequence was synthesized using the Oligo Design toolbox from Integrated DNA Technologies in order to be included as a transfection reagent, allowing part of the transfected clones to incorporate this template when repairing the cut by homologous recombination.

2.4.1.5. Miniprep

The newly generated vector expressing gRNA and Cas9 was purified following the instructions of SPEEDTOOLS Plasmid DNA purification kit (Biotools, 21.222) and sequenced to confirm that the gRNAs were correctly cloned.

2.4.1.6. Transfection of mESC

mESC were grown on 1% gelatin coated 6-well plates with standard mESC growth media. Once they reached total confluence, they were split 1:6 to get the optimal confluence for transfection. The day after the splitting, their transfection was performed using Lipofectamine ® 3000 transfection reagent (Thermo Fisher Scientific,) according to manufacturer's instructions. The transfection was performed with 500 ng of the vector pX330A_hCas9_long_chimeric_gRNA_G2P (containing the sgRNA sequence, the Cas9 coding sequence, GFP sequence and puromycin resistance) and three different amounts of the repairing template (a. 1ng, b. 500 ng., c. 100ng). PCR (primers in Table 8) of the three transfected heterogeneous populations (a, b and c) confirmed that the optimal transfection had been achieved with condition c, so this was the cell population used for the following steps.

Transfection efficiency was checked by GFP signal 16-18 hours after the transfection. When transfection efficiency was between 40-70%, puromycin selection was performed for 48 hours at

2µg/ml concentration. Afterwards, surviving cells were isolated in 96-well plates by serial dilution and, after expansion; clones with the chosen insertions were identified by PCR. Using the primers listed in Table 8, 24 mESC clones were investigated by PCR to check, on the one hand, whether they had inserted the Flag-HA epitope; and on the other, if the resulting whole tagged-region was at the desired genomic site and had the expected size (see Fig.5.B).

Table 8. Primers used for genotyping the mESC clones after transfection.

Primer	Sequence
p1	TGGCTACGAGTCGCTCACAC
p2	GACGACGACGACAAATACCC
p3	GGGCTAGGGAAGTTTGGCTC

2.5. Polymerase Chain Reaction (PCR)

2.5.1. RT-qPCR

In order to determine the transcripts levels of our cells, RNA was extracted and cDNA was produced as described previously (2.2.4. cDNA generation). Prior to RT-qPCR, samples were diluted 1:3 in dH_2O and 0.4µl of this dilution was used per 10µl reaction, according to the following Master Mix:

Mix RT-qPCR reaction. 1X (10µl)

5µl NZYSpeedy qPCR Green Master Mix (2X), ROX plus (NZY Tech)

0.125µl 10µM F, R primer Mix

0,4µl cDNA dilution

4.475µl H_2O

Table 9. Amplification program used for RT-qPCR reactions.

Step	Temperature	Time
Initial Denaturation	50.0	2 min
	95.0	2 min
40 cycles	95.0	15 sec
	60.0	15 sec
	72.0	1 min
Melting curve	95.0	15 sec
	60.0	1 min
	95.0	15 sec

RT-qPCR reactions were performed in 96 well plates on the StepOne Plus Thermocycler (Applied Biosystems) with the thermocycling program described in Table 9, using *Eef1α1* and *Hprt1* as housekeeping genes. Analysis of the resulting amplification curves was performed with the second derivative maximum method. All measurements were performed as triplicates and standard deviations were represented as error bars. The primers used for the expression analysis of the target genes are shown in Table 10.

Table 10. Primers used for expression analysis by RT-qPCR.

Primer	Sequence
Eef1a1_cDNA_F	TAGACGAGGCAATGTTGCTG
Eef1a1_cDNA_R	AGCGTAGCCAGCACTGATTT
Sox1_cDNA_F	CAACCAGGATCGGGTCAA
Sox1_cDNA_R	GTTTTCCTGGGCCATCTTG
Lhx5_cDNA_F	CGGGAAGGCAAGCTATACTG
Lhx5_cDNA_R	CAGGTCGCTCGGAGAGATAC
Zfp42_cDNA_F	GCGGTGTGTACTGTGGTGTC
Zfp42_cDNA_R	GACAAGCATGTGCTTCCTCA
Hprt1_cDNA_F	CAAGGGCATATCCAACAACA
Hprt1_cDNA_R	GCCCCAAAATGGTTAAGGTT
Six3_cDNA_F	CCTCACCCCCACACAAGTAG
Six3_cDNA_R	CTGATGCTGGAGCCTGTTCT
Lmx1b_cDNA_F	GGGATCGGAAACTGTACTGC
Lmx1b_cDNA_R	GTAGGGGCGATCTTCTCCAT
Lmx1a_cDNA_F	GAGACCACCTGCTTCTACCG
Lmx1a_cDNA_R	ACGGATGACAAACTCATTGG
Zic2_cDNA_F	CCGAGAACCTCAAGATCCAC
Zic2_cDNA_R	TGCATGTGCTTCTTCCTGTC
Nkx6.1_cDNA_F	CACGCTTGGCCTATTCTCTG
Nkx6.1_cDNA_R	GCGTGCTTCTTTCTCCACTT

2.5.2. ChIP-qPCR

ChIP DNA samples were analyzed by q-PCR to detect the level of enrichment of DNA polymerase II, H3K4me3 histone mark and ZIC2. Input samples were diluted 1:10, histone ChIP samples were diluted 1:3, and the rest of ChIP samples were used undiluted prior to qPCR. PCR reactions were performed in 96 well plates on the StepOnePlus Real-Time PCR device (Applied Biosystems) following the thermocycling program described in Table 10, in a volume of 10µl per reaction and using the same master as in 2.5.1. RT-qPCR. ChIP-qPCR signals were calculated as percentage of input. Standard deviations were measured from the technical triplicate reactions. The average of the percentage of input calculated for each ChIP sample was normalized to the signal average obtained in the same sample when two different negative control regions were used (Chr 2 (-), Chr 6 (-)). Primers used for the qPCRs of ChIP samples are listed in Table 11.

Table 11. Primers used for ChIP-qPCR.

Primer	Sequence	Product size
Chr6_neg_F	CTGGACTGAGGACCTTCTGC	161
Chr6_neg_R	AGGAAGGCAGATGAGGGATT	
Chr2_neg_F	CCTGAGGCTGGAAGTTTCTG	109
Chr2_neg_R	CTCCTGGGATTAAAGGCACA	
Lmx1a_pos_ChIP_Zic2_F	GAGCCTAGGGTGGGAATCTC	100
Lmx1a_pos_ChIP_Zic2_R	AGAAAGCCACTGGTGACTGC	
Lmx1b_pos_ChIP_Zic2_F	CGCTAGAGCCGCTTAATCAC	102
Lmx1b_pos_ChIP_Zic2_R	TCTGTACCTGCTGGGAGCTT	
Wnt3a_pos_ChIP_Zic2_F	CCATAGGCTGAGCACACAGA	151
Wnt3a_pos_ChIP_Zic2_R	AACTCACCCCAAGCCCTACT	
Nanog_pos_ChIP_Zic2_F	CTAGAGATCGCCAGGGTCTG	80
Nanog_pos_ChIP_Zic2_R	CCCCAAAAAGAGGCTTTACC	
Prmd4_pos_ChIP_Zic2_F	CATGGCCTCAAGTAGGGAAG	100
Prmd4_pos_ChIP_Zic2_R	TCGAAACCCGATTACTCCTG	
Eef1a_ChIP_pos_F	ACTCCTGTCCCTCCATTCCT	95
Eef1a_ChIP_pos_R	CGTTCAGCAATAGGGGCTAA	
Zfp42_ChIP_pos_F	AGCACACGAACACTTGGAAC	127
Zfp42_ChIP_pos_F	GTCCATTGGCCATCACGTTT	

2.5.3. Colony-PCR

To assess whether the transfected bacterial colonies carried the plasmid of interest (pX330A_hCas9_long_chimeric_gRNA_G2P) with the sgRNA insertion, PCRs were performed using OneTaq DNA polymerase (New England Biolabs). A selection of 7 bacterial colonies were resuspended in 30µl of dH_2O and heated at 95° to lysate the cells. 3µl of the lysate were used per PCR reaction. To amplify the region of interest, an oligo binding to the sequence of the gRNA

inserted (F: 5'-CACCCAATGAATGGTACGTGTGAG-3') was used as forward primer and one binding to a region of the vector adjacent to the insertion was used as reverse primer (R: 5' GGAAAGTCCCTATTGGCGTT-3'). If the gRNA insert was present in the vector, a band of 288 bp was expected (data not shown).

2.5.4. Clonal genotyping PCR

To define the genotype of the Flag-HA clones obtained after transfection, DNA was isolated using the protocol described in previous section 2.2.1. Genomic DNA isolation, and PCR was subsequently performed using OneTaq polymerase with GC Rich Buffer 5X, due to the CG rich sequence of the region of interest.

Mix OneTaq genotyping reactions.1X (25µl reaction)

5X GC Buffer	5µl
10 mM dNTPs	0,25µl
10 µM forward primer	0,5 µl
10 µM reverse primer	0,5 µl
One Taq Polymerase	0,125 µl
Template DNA	1µl
Nuclease free water	12,63µl

2.6. Statistical Analysis

2.6.1. RT-qPCR statistical analysis

The $2^{\Delta CT}$ Method was applied to calculate relative gene expression levels. C_T is defined as the fractional PCR cycle number at which the reporter fluorescence is greater than the threshold[26]. For each technical triplicate, the following calculations were performed:

1. $CT\ gene\ of\ interest - CT\ housekeeping\ genes = \Delta CT$
2. $Relative\ expression\ level = 2^{\Delta CT}$

Next, average relative expression level and standard deviations were calculated. Standard deviations were plotted as error bars.

2.6.2. ChIP-qPCR statistical analysis

Technical triplicates were used to calculate ChIP signals as percentage of input. Then, fold enrichments were calculated dividing the percentage of input of each region of interest by the average percent input of the negative regions (Chr2 and Chr6). For each technical triplicate:

1. $Input\ percentage = 100 \times 2^{(Average\ Input\ CT - Sample\ CT)}$
2. $Fold\ enrichment = \frac{Input\ percentage\ region\ of\ interest}{Average\ input\ percentage\ negative\ regions}$

Then, standard deviations were calculated from technical triplicate reactions and represented as error bars.

3. Results

3.1. Validation of gene expression differences between WT and *Zic2*$^{-/-}$ cells upon differentiation of mESC into AntNPCs.

The *in vitro* differentiation protocol whereby mESC are differentiated into AntNPCs (2.1.4. Differentiation of mESC into AntNPCs) partially recapitulates the early stages of brain development. As a result, after five days of differentiation (D5), cells with mostly an anterior and dorsal neural progenitor identity are obtained (i.e. AntNPC), as exemplified by the strong induction of *Lhx5* and *Lmx1b*, respectively.

To start elucidating the molecular mechanisms by which ZIC2 contributes to proper brain development, we previously analyzed by RNA-seq how the loss of ZIC2 function affected the differentiation of ESC into AntNPCs. Then, a differential gene expression analysis of this RNA-seq data revealed that the loss of ZIC2 disrupted dorsoventral patterning in AntNPC. To validate these previous results, I differentiated WT and *Zic2*$^{-/-}$ mESC lines into AntNPCs and then measured the expression levels of selected genes (based on the previous RNA-seq analysis) by RT-qPCR (Fig.4). Upon differentiation, both cell lines similarly downregulated the pluripotency marker *Zfp24*, which indicates a successful exit of pluripotency and lineage commitment (Fig. 4.A). As expected, the knock-out cell line is deployed of *Zic2* expression (Fig 4.B). Notably, *Zic2* knock out led to a drastic downregulation of dorsal brain genes, in particular, the classical roof plate markers *Lmx1a* and *Lmx1b (Fig*. 4. C and D). This downregulation of dorsal fates was coupled to an upregulation of ventral genes, as exemplified by the upregulation of the ventral marker *Nkx6.1* in *Zic2*$^{-/-}$ versus WT (Fig 4. E). In addition, *Lhx5*, a marker of the choroid plexus at more advanced stages of neural development, was upregulated in the knock out (Fig. 4. F), consistent with the phenotype observed in *Lmx1a*$^{-/-}$ / *Lmx1b* $^{-/-}$ mice [27]. Overall, these gene expression differences between WT and *Zic2*$^{-/-}$ AntNPC are in perfect agreement with our previous RNA-seq results, strongly indicating that the loss of ZIC2 impairs the establishment of brain dorsal identities, particularly the roof plate.

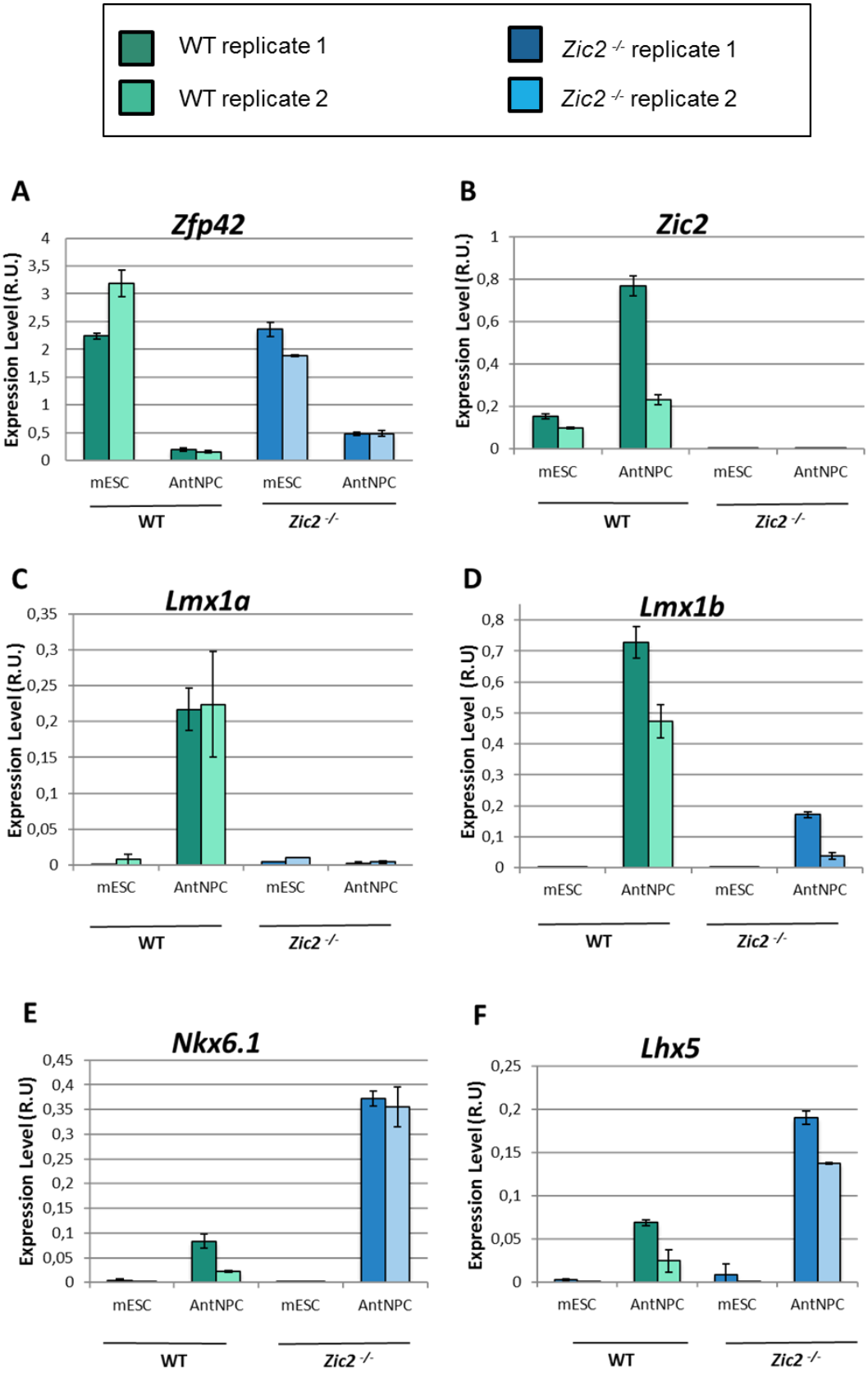

Figure 4. *Zic2* knock-out leads to an upregulation of dorsal genes and a downregulation of ventral ones in AntNPCs. The expression of (**A**) pluripotency marker *Zfp42*, (**B**) *Zic2*, (**C, D, F**) dorsal identity genes and (**E**) ventral identity markers was measured by RT-qPCR. For each gene, the results obtained using WT mESC and Zic2$^{-/-}$ mESC in two biological replicates are shown. Expression values were normalized to two housekeeping genes (*Eef1a1* and *Hprt1*) and the error bars represent standard deviations from three technical replicates. R.U.: Relative Units.

3.2. Generation of a *Zic2*-Flag-HA tagged mESC line using CRISPR-Cas9 technology

Next, we wanted to determine if ZIC2 directly activates these dorsal genes during AntNPC differentiation or if, alternatively, it represses ventral regulators which themselves can antagonize brain dorsal identity. To distinguish between these two possibilities, it is essential to globally identify the ZIC2 binding sites in AntNPC. To do so, our team previously performed ChIP for ZIC2 using a commercial ZIC2 antibody. However, IF and ChIP experiments on WT versus *Zic2*$^{-/-}$ cells demonstrated that the antibody was not completely specific for ZIC2. To overcome this limitation, we decided to generate a mESC line in which with the endogenous *Zic2* gene was tagged with a C-terminal Flag-HA epitope. To do so, we followed a CRISPR-Cas9 strategy based on two main steps: firstly, introduction of a double strand break close to *Zic2* transcriptional stop codon; secondly, homology induced repair (HDR) by the incorporation of a repair template containing the Flag-HA sequence (Fig. 5.A).

After transfection, single cells were plated in a 96 well plate in order to derive clonal lines. In total, 24 clones were viable. Genotyping by PCR showed that 12 of these clones were homozygous for the Flag-HA insertion: first, a PCR product of the expected size (334 bp) was obtained when using a forward primer that only hybridized to the Flag-HA tag region (p2) (Fig. 5.B); secondly, a single PCR product of the expected size (650 bp) was obtained when using primers flanking the inserted tag (p1-p3) (Fig.5.B).

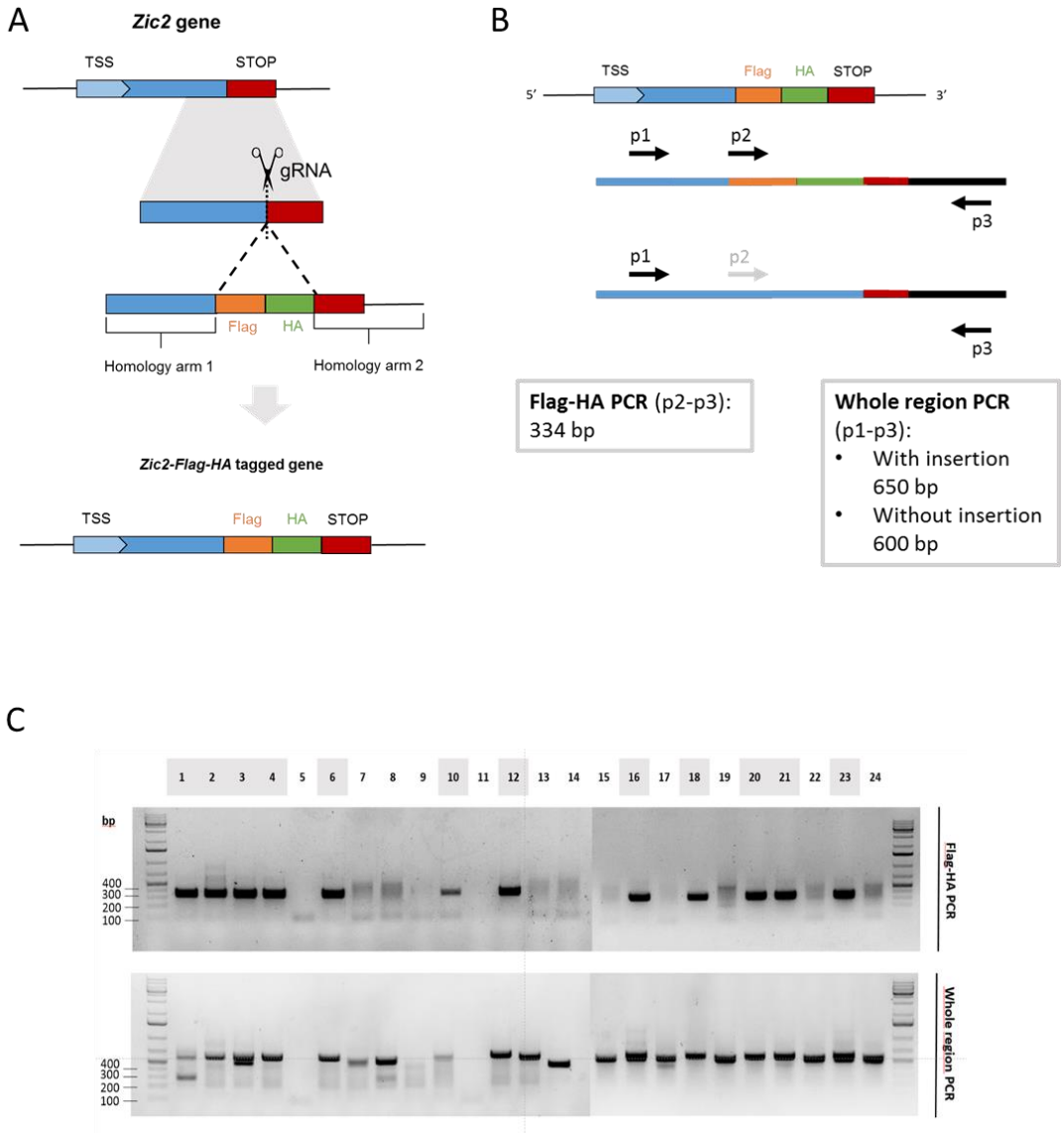

Figure 5. CRISPR-Cas9 strategy and genotyping. A. gRNAs were designed to target Cas9 and introduce a double-strand DNA break just before the stop codon. In addition, an oligonucleotide containing two homology arms and the Flag-HA sequence was also designed to be used as a repair template for homologous directed repair. **B.** For genotyping, two PCRs were performed for each clone. "Flag-HA PCR" were performed with primers p2 (hybridizing to the Flag tag) and p3 (3' of the Flag tag). "Whole region PCR" were performed with primers p1 and p3, which were flanking the Flag-HA insertion. **C.** PCR genotyping results of the 24 clones isolated after transfection. Marked in grey the clones selected for sequencing.

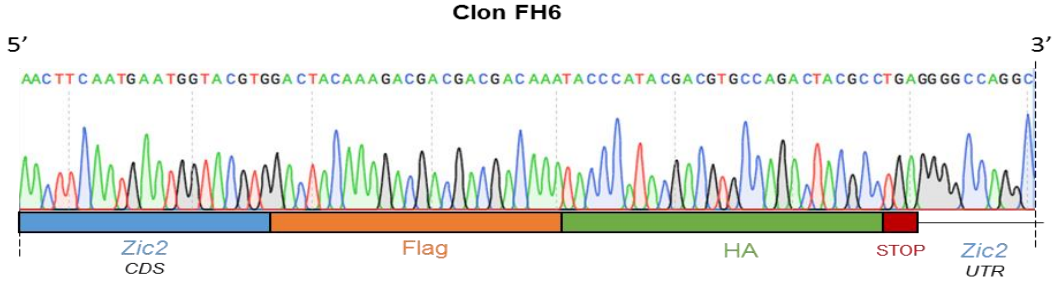

Figure 6. Illustrative example of the sequencing results obtained for the FH6 clonal mESC. CDS: coding sequence. UTR: 3' untranslated region.

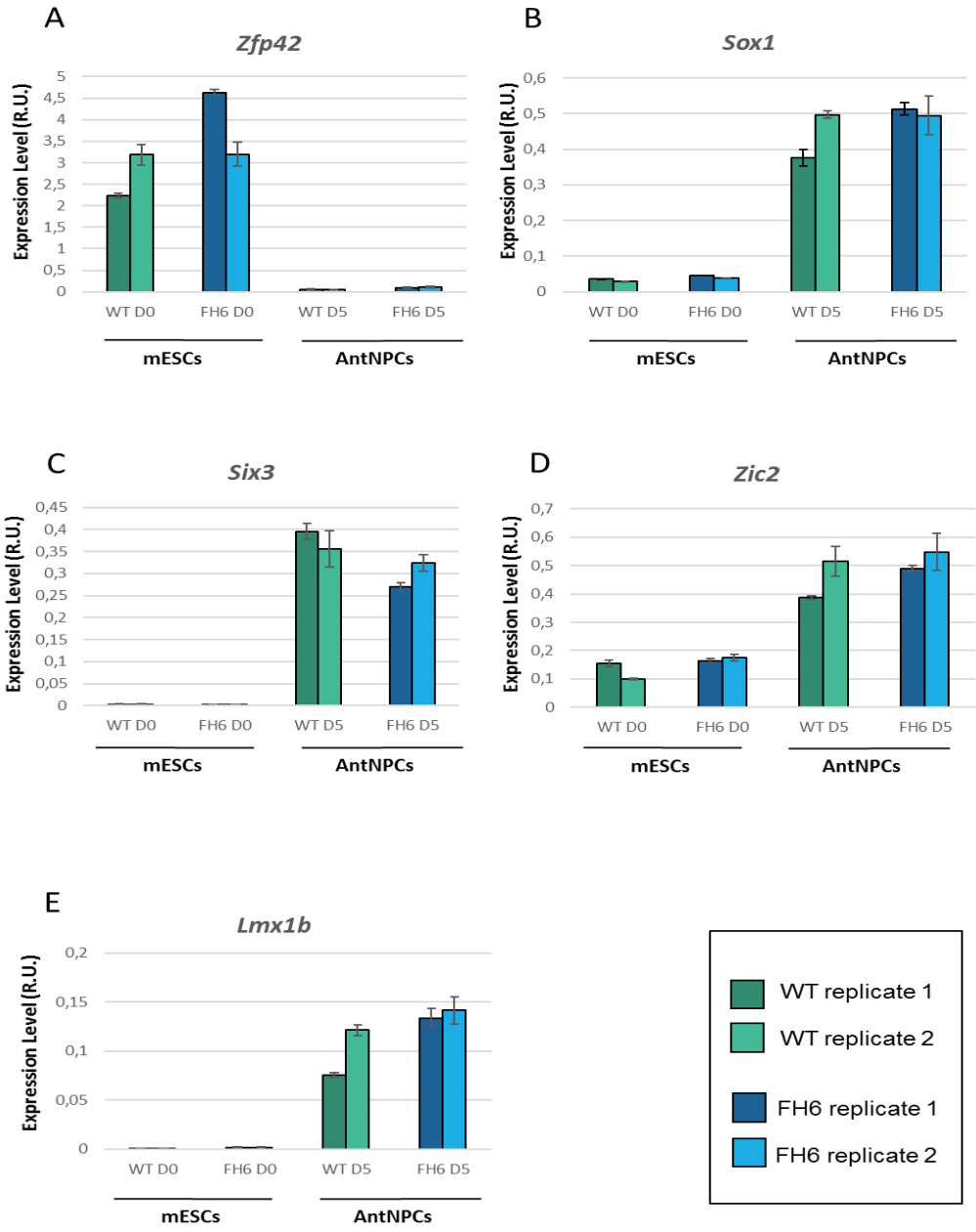

Figure 7. The generated FH6 mESC line differentiates as a WT into AntNPCs. The expression of (**A**) pluripotency marker *Zfp42* (**B, C, E**) neural identity genes *Sox1, Six3* and *Lmx1b* and (**D**) *Zic2* was measured by RT-qPCR. For each gene, the results obtained using WT mESC and FH6 mESCs in two biological replicates are shown. Expression values were normalized to two housekeeping genes (*Eef1a1* and *Hprt1*) and the error bars represent standard deviations from three technical replicates. R.U.: Relative Units.

The PCR products obtained with primers p1 and p3 ("Whole region PCR") (Fig.5.C) for the clones considered as homozygous for the Flag-HA insertion were sequenced by Sanger sequencing. The results confirmed that at least two of the clones, FH6 and FH12, were homozygous for the insertion and had the Flag-HA tag properly inserted just before the *Zic2* stop codon without any alterations on the *Zic2* open reading frame (Fig.6).

Based on the previous results, we then decided to more thoroughly characterize the FH6 mESC clonal line. Firstly, this mESC line was differentiated into AntNPCs in parallel with its parental WT mESC. The expression levels of important genes were measured by RT-qPCR. Both WT and FH6 cell lines differentiated properly and the FH6 line did not show any obvious morphological or proliferative differences in comparison to WT cells. Both FH6 and WT cells showed a downregulation of *Zfp42* upon differentiation, indicating loss of pluripotency (Fig.7.A). Neural markers, such as *Six3* or *Sox1*, were similarly upregulated upon differentiation (Fig. 7. B, C). Importantly, the expression levels of *Zic2* were similar in WT and FH6 lines, indicating that the tag was not affecting the proper expression of *Zic2*.

To confirm whether the normal *Zic2* expression levels observed by RT-qPCR in the FH6 cells could also be detected at the protein level, we performed Western Blot (WB) experiments. Firstly, we used the previously mentioned ZIC2 commercial antibody because previous work in our team demonstrated that, despite its lack of specificity in either IF or ChIP, this antibody was specific for ZIC2 in WB. These experiments again confirmed that ZIC2 protein levels were similar in WT and the tagged mESC. Next, I tried to detect ZIC2 with an anti-Flag antibody and I could not detect any signal, probably due steric effects that buried the Flag-HA epitope. Nevertheless, I then tried with an anti-HA antibody and, importantly, I could detect the ZIC2-Flag-Ha protein only in FH6 and FH12 mESC. This confirms that in FH6 mESC ZIC2 is stably expressed as a fusion protein containing the C-terminal Flag-HA tag.

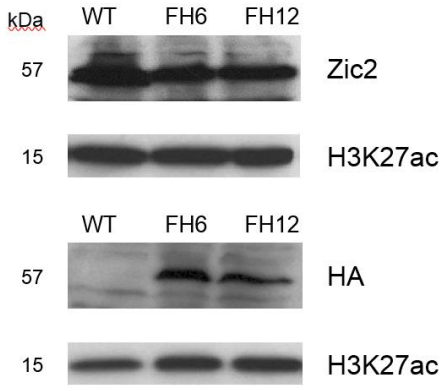

Figure 8. Flag-HA tagged ZIC2 levels in FH6 mESC are similar to endogenous ZIC2 levels in WT mESC. ZIC2 was detectable in WT and FH lines using a commercial antibody (Zic2), but the HA epitope was only detectable in the tagged mESC lines. The histone mark H3K27ac is shown as a loading control.

3.3. Optimization of sonication conditions for ChIP experiments

The newly generated *Zic2*-Flag-HA mESC line could in principle be used to perform ZIC2 ChIP experiments. To start evaluating whether this was the case, we first optimized sonication conditions on WT mESC lines. For optimal ChIP when working with TFs, the size of the

sonicated DNA fragments should be in a range of 200 to 1000 bp. To achieve these conditions, I tried two different approaches.

First, we sonicated the mESC with the Bioruptor Plus sonication device (Diagenode) for a total of 45 cycles (30 seconds on/ 30 seconds off) taking aliquots every 5 cycles. After running the sonicated DNA on an agarose gel we observed that, with this sonicator, the best results were obtained after 45 cycles (Fig 9.A). However, even under this conditions, many of the resulting DNA fragments were still too large (>2Kb) while others were too small (<200 bp). 45 minutes is an extensive sonication time, which could somehow compromise the chromatin integrity due excessive warming of the samples. Hence, we tried with another sonication device.

mESC were then sonicated with Ultrasonic processor (Labsonic. Parameters: Amplitude 80%) (Fig.9. B). In this case, 10 sonication cycles (30 seconds on/ 30 seconds off) resulted in DNA fragments that were really close to the desired size distribution, whilst 20 cycles already resulted in excessive fragmentation (Fig. 9. B). Finally, 13 cycles were chosen as the optimal sonication conditions for this device and were used in subsequent ChIP experiments.

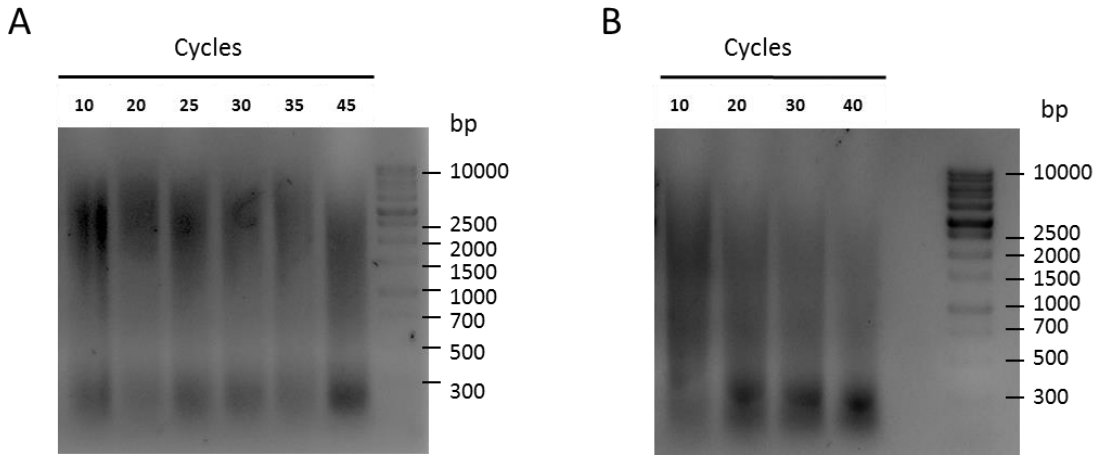

Figure 9. Sonication optimization. Agarose gel electrophoresis of the sonicated mESC samples after an increasing number of cycles using **(A)** Bioruptor Plus and **(B)** Ultrasonic processor.

3.4. Testing ChIP with Pol2 and H3K4me3 in control regions

Once sonication conditions were optimized for mESC, we then wanted to ensure that whole ChIP protocol was working in my hands. Therefore, I performed ChIP-qPCR in WT cell lines for DNA Polymerase II (Pol2) and for the histone mark H3K4me3. Based on publically available ChIP-seq data in mESC, we designed primers to amplify two negative and two positive control regions for the binding of Pol2 and the presence of H3K4me3 (Fig. 10). qPCR analyses of the Pol2 and H3K4me3 ChIP samples confirmed strong fold enrichments for both Pol2 (Fig. 10.A) and for H3K4me3 (Fig. 10.B) in the positive regions compared to the negative ones.

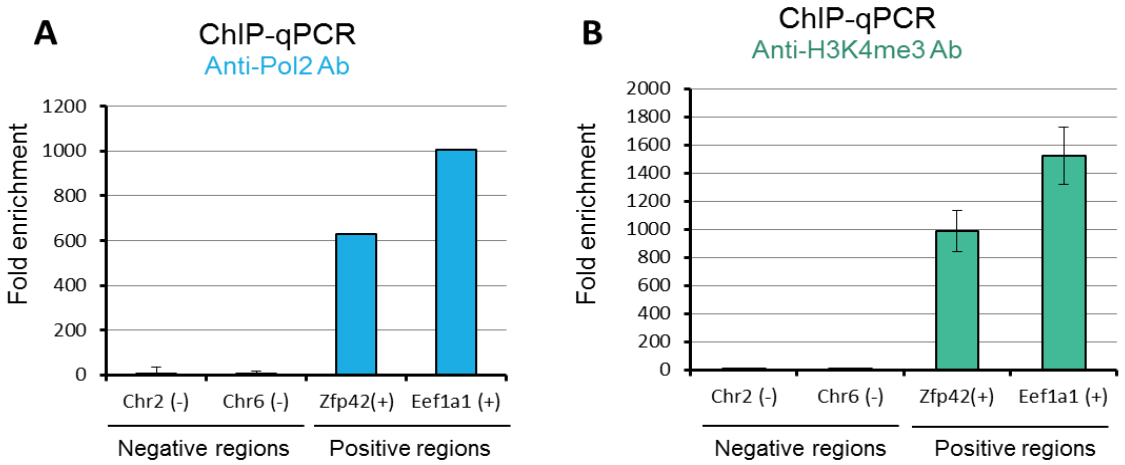

Figure 10. Testing ChIP conditions in WT mESCs for positive and negative control regions. ChIP-qPCR was performed to evaluate the binding of (**A**) DNA Polymerase II (Pol2) and the presence of (**B**)H3K4me3 in two positive control regions in *Zfp42* and *Eef1a1* promoters and in two negative control regions in chromosome 2 and chromosome 6. The qPCRs were performed as technical triplicates. For each technical triplicate, fold enrichments were calculated as the %input of the region analyzed divided by the average % input of the two negative control regions (chr2 and chr6). Standard deviation values are plotted as error bars.

3.5. ZIC2 ChIP validation in control negative regions and putative positive binding regions

Next, we wanted to determine if the tagged FH6 mESC line was suitable to identify ZIC2 binding sites by ChIP and if the antibodies for the two Flag and HA epitopes could be used for this purpose. To do so, based on the ChIP-Seq data on mESCs published by Luo et al. 2015, we selected some control negative regions in chromosome 2 and chromosome 6 where no ZIC2 binding was expected and some putative positive regions for ZIC2 binding (Fig 11).

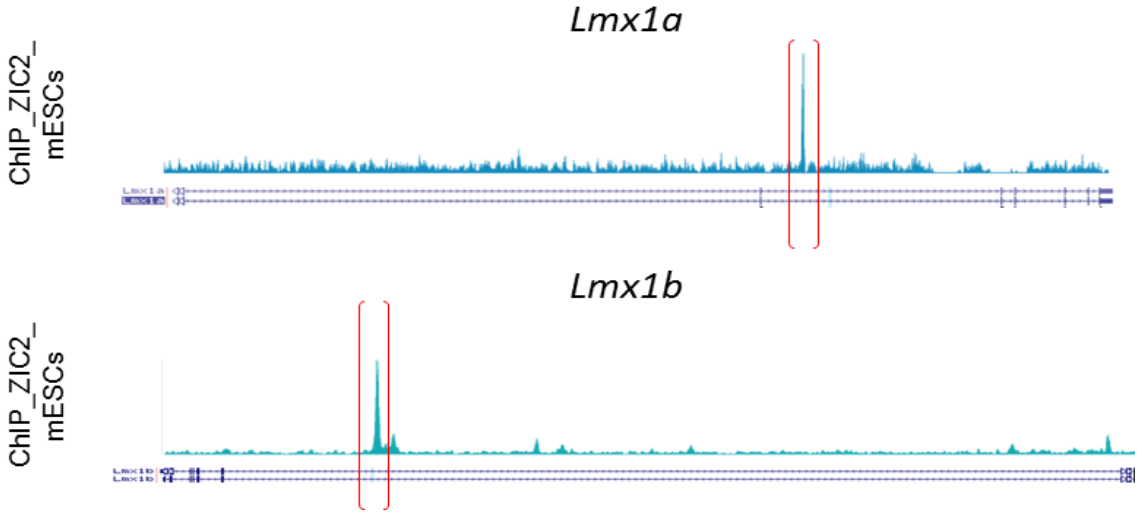

Figure 11. Example of two of the five selected positive regions for testing ZIC2 ChIP in FH6 mESC. Diagram showing ZIC2 binding obtained from Luo et al. 2015 data wherein the regions amplified by RT-qPCR in our ChIP are marked between red brackets. Whole gene regions obtained from UCSC genome browser.

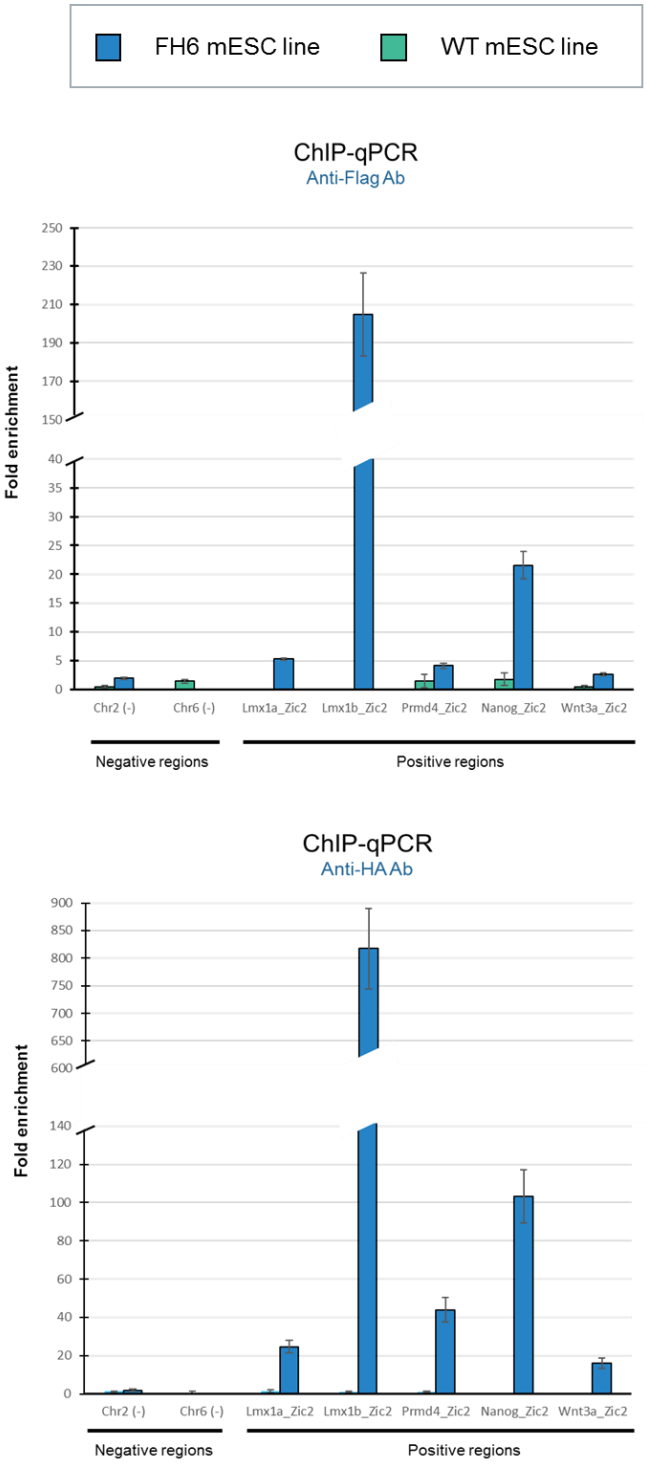

Figure 12. The FH6 mESC line is suitable to investigate ZIC2 binding profiles by ChIP. ChIP-qPCR experiments were performed for the positive and negative ZIC2 binding regions in WT and FH6 mESC lines, using either the anti-HA (A) or anti-Flag antibodies (B). The qPCRs were performed as technical triplicates. For each technical triplicate, fold enrichments were calculated as the %input of the region analyzed divided by the average % input of the two negative control regions (chr2 and chr6). Standard deviation values are plotted as error bars.

ChIPs using antibodies against FLAG and HA were performed in both WT and in FH6 mESC. Then, the resulting ChIP DNA samples were analyzed by qPCR. Notably, positive regions showed much higher enrichments levels in FH6 mESC than in WT mESC for both the FLAG and HA antibodies. However, when more carefully comparing the enrichments obtained for the two antibodies, it seems that the anti-Flag gives lower fold-enrichments than the anti-HA antibody. Thus, the anti-HA antibody would be preferable for future ZIC2 ChIP experiments. Overall, these ChIP-qPCR experiments confirm that ZIC2 binds to the evaluated regions in mESC and also demonstrate that the FH6 cell lines can be used to investigate ZIC2 binding profiles by ChIP-seq.

4. Discussion

The comparison of the expression levels of selected genes between *Zic2*$^{-/-}$ and WT AntNPC by RT-qPCR is in perfect agreement with our previous RNA-seq data. Altogether, this clearly confirms that loss of ZIC2 leads to a general downregulation of dorsal identity genes and to an upregulation of ventral ones. Remarkably, the expression of the classical roof plate markers *Lmx1a* and *Lmx1b* is almost completely lost in *Zic2*$^{-/-}$ cells upon differentiation. These results recapitulate the brain dorsal patterning defects previously observed in mouse embryos deficient for *Zic2* [28].In contrast to the report from Luo et al.[29], but in agreement with previous *in vivo* observations, our *Zic2*$^{-/-}$ mESC showed no obvious defects in self-renewal or pluripotency and they were able to differentiate into AntNPC without any major cell death. It would be interesting to analyze in more detail the transcriptional changes that occur during the differentiation of *Zic2*$^{-/-}$ mESC into AntNPC by generating RNA-seq data at additional time points (i.e. D0, D3, D4 and D5). This should help us to better define the differentiation stage/s at which gene expression defects arise and to relate such defects with other genomic data obtained at the same time points (e.g. ChIP-seq for ZIC2 and histone modifications). However, the current data already supports that ZIC2 plays a prominent role in the establishment of dorsal neural identities and, in particular, in roof plate induction. This is in perfect agreement with previous hypothesis regarding the etiology of MIHV HPE, according to which ZIC2 act as a roof plate inducer required for proper dorsal patterning of the brain (Cheng et al., 2006). What remains to be clarified is whether ZIC2 is acting as direct activator of dorsal genes, as a repressor of ventral ones or both. Given its diverse functions at different stages of neural development (Houtmeyers et al., 2013; Ali et al., 2012; Herrera et al., 2003; Escalante et al. 2007), it is tempting to speculate that ZIC2 might use different regulatory mechanisms depending on the cellular and genomic context. Therefore, different approaches should be ideally combined to uncover the regulatory networks and regulatory mechanisms by which ZIC2 controls early brain patterning.

To elucidate the gene regulatory networks directly controlled by ZIC2 upon AntNPC differentiation, it will be important to uncover the ZIC2 binding sites during AntNPC differentiation. This would enable us to compare the ZIC2 binding profiles with the transcriptional changes observed upon loss of ZIC2 in order to identify the genes that are directly controlled by ZIC2. Moreover, these comparisons should also reveal whether ZIC2 preferentially activates or represses its direct target genes. Previous work in our team using a ZIC2 commercial antibody and *Zic2*$^{-/-}$ mESC (same used in Luo et al. 2015) showed that, although the antibody was specific according to WB analysis, it was not for either IF or ChIP. To solve this issue, in this project I generated a *Zic2*-Flag-HA tagged mESC line using CRISPR-Cas9 technology. To confirm that the addition of the Flag-HA epitopes did not affect the expression and/or function of ZIC2 I performed several experiments. First, I confirmed that the expression of *Zic2* was not altered at the mRNA levels based on RT-qPCR analyses. Moreover, the tagged mESC line behaved similarly to its parental WT mESC, with no major alterations in expression, morphology or proliferation. Next, I tried to detect tagged ZIC2 protein using antibodies against either Flag or HA. Initially, I faced problems detecting the Flag epitope by WB in the tagged mESC. I tried two different primary anti-Flag antibodies without success. This could be explained by the conditions used during WB transfer that can lead to changes in the folding of the tagged protein and difficulties in detecting the Flag epitope due to steric effects. Nevertheless, these problems were not observed when using an anti-HA antibody, which successfully detected the ZIC2 protein only in the tagged mESC lines. In contrast to the denaturing conditions used in WB, in ChIP proteins are believed to maintain their native 3D conformation. Therefore, I decided to perform ChIP in the ZIC2-tagged cell line using both Flag and HA antibodies. The results demonstrate that both antibodies can be used to identify ZIC2 binding sites. However, when comparing to ChIPs performed in WT cells in which the Flag-HA epitopes are not present and that, consequently, represent an excellent negative control, I observed that the HA antibody was more specific and yielded higher fold-enrichment. Therefore, I conclude, that the HA antibody is preferable for ChIP and should be ideally used in future ChIP-seq experiments.

Once I confirmed the gene expression changes previously observed in *Zic2*$^{-/-}$ AntNPC and having established a *Zic2-FH* mESC line suitable for ChIP, we can now generate additional RNA-seq transcriptional profiles in WT vs *Zic2*$^{-/-}$ at different time points of the AntNPC differentiation (i.e. D0, D3, D5) and coupled this to ZIC2 ChIP-seq data that can be generated at similar differentiation stages. This should help us to determine which are the direct ZIC2 target genes and whether ZIC2 preferentially acts as an activator, a repressor or both. Moreover, we could also perform ChIP-seq for different active and repressive histone marks in WT vs *Zic2*$^{-/-}$ during AntNPC differentiation that will inform us about the chromatin changes that occur at both ZIC2 binding regions and its target genes. This will allow us to understand how ZIC2 binding affects

the chromatin state of its genomic targets and of the genes it regulates. Most likely this will shed some light into the role of ZIC2 in the activation of poised enhancers during the differentiation of mESC into AntNPCs.

Once we characterize the genomic binding dynamics and the transcriptional networks controlled by ZIC2 another interesting aspect of the project could be to uncover the trans-regulatory mechanism/s by which ZIC2 controls the expression of its direct target genes. To do so, two main research lines could be followed. On the one hand and using the recently generated Zic2-Flag-HA tagged cell line, we could perform ZIC2 IP coupled to MS experiments to identify ZIC2 interacting partners such as co-activators or co-repressors. On the other hand, given the phosphorylation-dependent transactivation activity of ZIC2 [30,31] we could use CRISPR-Cas9 technology to mutate the phosphorylation site of ZIC2 and analyze possible changes in expression in comparison to WT cell lines. These experiments could establish whether ZIC2 function upon differentiation into AntNPCs is strictly dependent on its phoshorylation. If this is the case, it would also be highly relevant to uncover the kinase and signaling pathway responsible of phosphorylating ZIC2 during AntNPC phosphorylation.

Uncovering the regulatory networks and mechanisms whereby ZIC2 contributes to brain development are essential steps towards elucidating the etiology of ZIC2-associated HPE. As mentioned in the introduction, ZIC2-associated HPE is characterized by its reduced penetrance and variable expressivity, which strongly suggests that, together with loss of ZIC2 function, other genetic and/or environmental risk factors affect this disorder. In addition, human and mice seem to display different ZIC2 dosage sensitivities, as ZIC2 heterozygosis can lead to HPE in humans while not causing major abnormalities in mice [28,32]. We hypothesize that the reduced penetrance of ZIC2-dependent HPE and the differential ZIC2 dosage sensitivity between mice and humans could be caused by unexplored genetic (*e.g.* ZIC5) and/or gene-environmental (GxE) interactions (*e.g.* exposure to Ethanol [17] or Retinoic Acid [33,34]), which could increase the transcriptional and phenotypic defects caused by low ZIC2 levels alone. To test the relevance of GxE interactions in the etiology of ZIC2-associated [25,33] HPE, we will use CRISPR/Cas9 technology to generate mouse and human ESC lines that are heterozygous for *ZIC2* and model the defects observed in HPE. Then, WT mESC and WT hESC (human embryonic stem cells) will be differentiated into AntNPC under increasing doses of ethanol or Retinoic Acid. Using the information generated in the previously mentioned RNA-Seq approach (data at D0, D3,D5) , expression levels of critical ZIC2 target genes will be measured by RT-qPCR in AntNPC derived from the different ESC genotypes and teratogenic treatments. According to our preliminary data in mouse *Zic2*$^{-/-}$ AntNPC, these target genes are likely to include major dorso-ventral patterning genes as well as neuronal differentiation markers. These initial gene expression analyses should reveal whether any of the investigated environmental interactions exists. Together with the time profile of RNA-

seq data upon differentiation that we plan to generate, these global analyses should conclusively show the extent to which ethanol and/or retinoic acid interact with *ZIC2* during anterior neural development in both mice and humans and, thus, potentially contribute to HPE aetiology.

Overall, my Master Thesis project represents an important starting point in order to globally and mechanistically unravel ZIC2 function as a transcriptional regulator during brain development. Notably, I have confirmed that (i) the the loss of ZIC2 disrupts dorsoventral patterning during the differentiation of mESC into AntNPCs and (i) I have generated a mESC line that can be used as an important tool to investigate ZIC2 function using various experimental approaches (ChIP, IP, IP-MS).

5. References

1. Houtmeyers R, Souopgui J, Tejpar S, Arkell R. The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. *Cell Mol Life Sci*. 2013;70(20):3791-3811. doi:10.1007/s00018-013-1285-5

2. Aruga J. The role of Zic genes in neural development. *Mol Cell Neurosci*. 2004;26(2):205-221. doi:10.1016/j.mcn.2004.01.004

3. Layden MJ, Meyer NP, Pang K, Seaver EC, Martindale MQ. Expression and phylogenetic analysis of the zic gene family in the evolution and development of metazoans. *Evodevo*. 2010;1(1):12. doi:10.1186/2041-9139-1-12

4. Ali RG, Bellchambers HM, Arkell RM. Zinc fingers of the cerebellum (Zic): Transcription factors and co-factors. *Int J Biochem Cell Biol*. 2012;44(11):2065-2068. doi:10.1016/j.biocel.2012.08.012

5. Herrera E, Brown L, Aruga J, et al. Zic2 patterns binocular vision by specifying the uncrossed retinal projection. *Cell*. 2003;114(5):545-557. doi:10.1016/S0092-8674(03)00684-6

6. Escalante A, Murillo B, Morenilla-Palao C, Klar A, Herrera E. Zic2-dependent axon midline avoidance controls the formation of major ipsilateral tracts in the CNS. *Neuron*. 2013;80(6):1392-1406. doi:10.1016/j.neuron.2013.10.007

7. Dubourg C, Bendavid C, Pasquier L, Henry C, Odent S, David V. Holoprosencephaly. *Orphanet J Rare Dis*. 2007;2(1):8. doi:10.1186/1750-1172-2-8

8. Cohen MM. Holoprosencephaly: Clinical, anatomic, and molecular dimensions. *Birth Defects Res Part A - Clin Mol Teratol*. 2006;76(9):658-673. doi:10.1002/bdra.20295

9. Fernandes M, Hébert JM. The ups and downs of holoprosencephaly: Dorsal versus ventral patterning forces. *Clin Genet*. 2008;73(5):413-423. doi:10.1111/j.1399-0004.2008.00994.x

10. Barratt KS, Arkell RM. ZIC2 in holoprosencephaly. In: *Advances in Experimental Medicine and Biology*. Vol 1046. ; 2018:269-299. doi:10.1007/978-981-10-7311-3_14

11. Lewis AJ, Simon ; E M, Barkovich ; A J, et al. *Articles Middle Interhemispheric Variant of Holoprosencephaly A Distinct Cliniconeuroradiologic Subtype*. Vol 59.; 2002. www.neurology.org. Accessed June 14, 2019.

12. Barkovich AJ, Quint DJ. Middle interhemispheric fusion: an unusual variant of holoprosencephaly. *AJNR Am J Neuroradiol*. 14(2):431-440. http://www.ncbi.nlm.nih.gov/pubmed/8456724. Accessed June 14, 2019.

13. Fernandes M, Gutin G, Alcorn H, McConnell SK, Hébert JM. Mutations in the BMP pathway in mice support the existence of two molecular classes of holoprosencephaly. *Development*. 2007;134(21):3789-3794. doi:10.1242/DEV.004325

14. Elms P, Siggers P, Napper D, Greenfield A, Arkell R. Zic2 is required for neural crest formation and hindbrain patterning during mouse development. *Dev Biol*. 2003;264(2):391-406. doi:10.1016/j.ydbio.2003.09.005

15. Nagai T, Aruga J, Minowa O, et al. *Zic2 Regulates the Kinetics of Neurulation.* Vol 97.; 2000. http://www.ncbi.nlm.nih.gov/pubmed/10677508%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC26484. Accessed September 18, 2018.

16. Petryk A, Graf D, Marcucio R. Holoprosencephaly: Signaling interactions between the brain and the face, the environment and the genes, and the phenotypic variability in animal models and humans. *Wiley Interdiscip Rev Dev Biol*. 2015;4(1):17-32. doi:10.1002/wdev.161

17. Cohen MM, Shiota K. Teratogenesis of holoprosencephaly. *Am J Med Genet*. 2002;109(1):1-15. http://www.ncbi.nlm.nih.gov/pubmed/11932986. Accessed June 17, 2019.

18. Brown L, Brown S. Zic2 is expressed in pluripotent cells in the blastocyst and adult brain expression overlaps with makers of neurogenesis. *Gene Expr Patterns*. 2009;9(1):43-49. doi:10.1016/J.GEP.2008.08.002

19. Elms P, Scurry A, Davies J, et al. Overlapping and distinct expression domains of Zic2 and Zic3 during mouse gastrulation. *Gene Expr Patterns*. 2004;4(5):505-511.

doi:10.1016/j.modgep.2004.03.003

20. Cruz-Molina S, Respuela P, Tebartz C, et al. PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell*. 2017;20(5):689-705.e9. doi:10.1016/j.stem.2017.02.004

21. Luo Z, Gao X, Lin C, et al. Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Mol Cell*. 2015;57(4):685-694. doi:10.1016/j.molcel.2015.01.007

22. Allen HF, Wade PA, Kutateladze TG. The NuRD architecture. *Cell Mol Life Sci*. 2013;70(19):3513. doi:10.1007/S00018-012-1256-2

23. Matsuda K, Kondoh H. Dkk1-dependent inhibition of Wnt signaling activates *Hesx1* expression through its 5′ enhancer and directs forebrain precursor development. *Genes to Cells*. 2014;19(5):374-385. doi:10.1111/gtc.12136

24. Charette SJ, Cosson P. Prearation of genomic DNA from Dictyostelium discoideum for PCR analysis. *Biotechniques*. 2004;36:574-575. https://pdfs.semanticscholar.org/6db0/62666f02339f426ce1d8c1f2ffcf2fc593af.pdf. Accessed June 15, 2019.

25. Boyer LA, Plath K, Zeitlinger J, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*. 2006;441(7091):349-353. doi:10.1038/nature04733

26. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. *Genome Res*. 1996;6(10):986-994. doi:10.1101/gr.6.10.986

27. Cheng X, Hsu C, Currle DS, Hu JS, Barkovich AJ, Monuki ES. Journal of Neuroscience. *J Neurosci*. 2006;20(7):2618-2625. doi:10.1523/jneurosci.0714-06.2006

28. Nagai T, Aruga J, Minowa O, et al. Zic2 regulates the kinetics of neurulation. *Proc Natl Acad Sci U S A*. 2000;97(4):1618-1623. http://www.ncbi.nlm.nih.gov/pubmed/10677508%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC26484. Accessed September 18, 2018.

29. Luo Z, Gao X, Lin C, et al. Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Mol Cell*. 2015;57(4):685-694. doi:10.1016/j.molcel.2015.01.007

30. Ishiguro A, Ideta M, Mikoshiba K, Chen DJ, Aruga J. ZIC2-dependent transcriptional regulation is mediated by DNA-dependent protein kinase, poly(ADP-ribose) polymerase, and RNA helicase A. *J Biol Chem*. 2007;282(13):9983-9995. doi:10.1074/jbc.M610821200

31. Ishiguro A, Aruga J. Functional role of Zic2 phosphorylation in transcriptional regulation. *FEBS Lett*. 2008;582(2):154-158. doi:10.1016/J.FEBSLET.2007.11.080

32. Warr N, Powles-Glover N, Chappell A, Robson J, Norris D, Arkell RM. Zic2-associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation. *Hum Mol Genet*. 2008;17(19):2986-2996. doi:10.1093/hmg/ddn197

33. Piersma AH, Hessel E V., Staal YC. Retinoic acid in developmental toxicology: Teratogen, morphogen and biomarker. *Reprod Toxicol*. 2017;72:53-61. doi:10.1016/j.reprotox.2017.05.014

34. Bauer LB, Ornelas JN, Elston DM, Alikhan A. Isotretinoin: controversies, facts, and recommendations. *Expert Rev Clin Pharmacol*. 2016;9(11):1435-1442. doi:10.1080/17512433.2016.1213629