

# Unraveling the function of trophectoderm markers in early embryonic development via CRISPR/Cas9 gene editing

Sarah Declercq

Student number: 01804130

Promotor: Prof. Dr. Björn Heindryckx

Co-promotor: Dr. Ir. Annekatrien Boel

Mentor: Gwenny Cosemans

Ghent Fertility and Stem Cell Team (G-FAST), Department for Reproductive  
Medicine, Ghent University Hospital, Corneel Heymanslaan 10, 9000 Ghent, Belgium

**A dissertation submitted to Ghent University in partial fulfillment of the requirements  
for the degree of Master of Science in Biomedical Sciences**

Academic year: 2022-2023

## Acknowledgements

The submission of this master's thesis concludes the amazing journey I have made to become a Master of Science in Biomedical Sciences. It has been a tough, hectic, beautiful, and educational part of my life, on which I will always look back with gratitude. I have grown as a scientist, but also as a person and I could not have done this without a multiple of people, to which I want to express great gratitude.

First, I want to thank my promotor, Prof. Dr. Björn Heindryckx, for handing me the opportunity to conduct the research for my master's thesis in the Ghent Fertility and Stem Cell Team and for correcting this master's dissertation. It has been a fascinating journey learning about not only my topic, but also other topics covered in the G-FAST lab. I also want to thank you for handing me the opportunity to continue this research in a PhD. This has been the opportunity of a lifetime for me. Furthermore, I also want to thank my co-promotor Dr. Ir. Annkatrien Boel, for analyzing the Miseq data, and for proofreading this thesis.

The person I want to throw the biggest thank you to is my mentor, Gwenny Cosemans. Ever since we started working together, I really looked up to you, and how passionate and motivated you were about your research. You are a positive, hard-working and kind-hearted person. Thank you for teaching me pretty much everything I know about our topic and all the techniques in the lab, and for the endless support you had for me. You taught me how to be an independent, critical thinking and hard-working scientist, which is something I always aspired to be. Of course, I also want to thank you for always rooting for me during my PhD application process, giving me the right advice and for helping me so much with the FWO application. I wish you so much luck in your further career, but I am sure everything you aspire to be, you will succeed in. I look forward to working together in the upcoming years and hope to exchange a lot of super nice results with each other!

Next, I would also want to thank all the other members of G-FAST for answering my questions, the reassuring words, or just for the many laughs and jokes during the (few) breaks we had together! A special thanks to Bieke Bekaert, for helping me so much, even though she was swamped with work herself. I wish you all the luck in your career, but I know you will kill it anyways. Also, I want to thank my fellow master thesis students, for all the laughs, support and stories we told each other in the student room.

Last but not least, I want to thank my parents and grandparents for their endless support, and for showing interest in my thesis work, even though they did not fully understand what was going on most of the time.

## List of abbreviations

<b>Abbreviation</b>	<b>Fully written</b>
AMOT	Angiomotin
AP2 $\alpha$	Activating enhancer binding protein 2 alpha
AP2 $\gamma$	Activating enhancer binding protein 2 gamma
ART	Assisted reproductive technology
B-domain	Basic domain
BMP2R	Bone morphogenetic protein 2 receptor
BMP4	Bone morphogenetic protein 4
BMP7	Bone morphogenetic protein 7
Bp	Base pair
BSA	Bovine serum albumin
Cas9	CRISPR associated protein 9
CDX2	Caudal Type Homeobox 2
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
CTB	Cytotrophoblast
DAPI	4',6-diamidino-2-phenylindole
DEPC	diethyl pyro carbonate
DKO	Double knock-out
DMEM	Dulbecco's modified Eagle medium
DNA	Deoxyribonucleic Acid
dpf	Days post fertilization
DSB	Double-stranded break
E	Embryonic day
EC	Ectoplacental cone
EPI	Epiblast
EVT	Extra villous trophoblast
FA	Fragment analyzer
FBS	Fetal bovine serum
FGF2R	Fibroblast growth factor 2 receptor
FGF4	Fibroblast growth factor 4
GATA2	GATA-binding factor 2
GATA3	GATA-binding factor 3
G/C	Guanine/cytosine
gRNA	Guide RNA
GOI	Gene of interest
GV	Germinal vesicle
hCG	Human chorionic gonadotropin
HDR	Homology-directed repair
hESC	Human embryonic stem cells
HSH	Helix-span-helix
ICM	Inner cell mass
ICSI	Inter cytoplasmic sperm injection
IGV	Integrative genomics viewer
IH-HR	Inter-homolog homologous repair
IP	Intraperitoneal
IF	immunofluorescent
IVF	<i>In vitro</i> fertilization
IVM	<i>In vitro</i> maturation
Kb	Kilo bases

KI	Knock-in
KO	Knock-out
LATS	Large tumor suppressor kinase
LIF	Leukemia inhibitory factor
LOH	Loss of heterozygosity
mESC	Mouse embryonic stem cells
MEM-NEAA	Minimal essential medium non-essential amino acids
MI	Metaphase 1
MII	Metaphase 2
mLIF	Mouse leukemia inhibitory factor
MMEJ	microhomology-mediated end-joining
mRNA	Messenger ribonucleic acid
NANOG	Nanog Homeobox
NHEJ	Non-homologous end joining
NMD	Nonsense-mediated decay
NRD	Negative regulatory domain
OCT4	Octamer-binding transcription factor 4
Opti-MEM	Opti minimal essential medium
ORF	Open reading frame
PBS	Primer binding site
PBS 1X	Phosphate buffered saline 1X
PBS 10X	Phosphate buffered saline 10X
P/Q	Proline/glutamine
PAM	protospacer adjacent motif
Par-aPKC	Atypical protein kinase C
PBS	Primer binding site
PCR	Polymerase chain reaction
pegRNA	Prime editing guide RNA
Pen strep	Penicillin streptomycin
PFA	Paraformaldehyde
PMSG	Pregnant mare serum gonadotropin
PRE	Primitive endoderm
Rcf	Relative centrifugal force
RhoA	Ras homolog family member A
RNA	Ribonucleic Acid
RNP	Ribonucleoprotein
RT	Room temperature
SER	Smooth endoplasmic reticulum
SMAD1	Suppressor of Mothers Against Decapentaplegic 1
SOX2	sex-determining region Y-box 2
STB	Syncytiotrophoblast
TAD	Trans-activating domain
TBX3	T-box transcription factor 3
TE	Trophectoderm
TEAD4	TEA domain transcription factor 4
TF	Transcription factor
TFAP2A	Transcription factor AP-2 alpha
TFAP2C	Transcription factor AP-2 gamma
TGC	Trophoblast giant cells
tracrRNA	trans-activating CRISPR RNA
YAP/TAZ	Yes-associated protein/ transcriptional coactivator with PDZ-binding motif
ZnF	Zinc finger
ZnFn	Zinc finger nuclease

## Contents

Acknowledgements .....	
List of abbreviations .....	
Summary.....	1
Societal impact.....	1
1. Introduction .....	2
1.1 Pre-implantation development.....	2
1.2 Peri-implantation development .....	3
1.3 Major genetic players during preimplantation development .....	4
1.3.1 Polarization.....	4
1.3.2 First lineage segregation.....	5
1.3.3 Second lineage segregation .....	7
1.3.4 Implantation .....	8
1.4 Interspecies differences .....	8
1.5 Genes of interest.....	11
1.5.1 <i>Tfap2a/TFAP2A, Tfap2c/TFAP2C</i> .....	11
1.5.2 <i>Gata2/GATA2, Gata3/GATA3</i> .....	13
1.6 CRISPR/Cas9 .....	15
1.6.1 Limitations .....	17
1.6.2 Alternative approaches .....	18
1.7 Scientific research objectives .....	20
2. Materials and methods .....	20
2.1 crRNA and primer design .....	20
2.2 Mouse zygote collection .....	21
2.3 RNP preparation for electroporation/nucleofection .....	21
2.4 Electroporation: CRISPR/Cas9 delivery in mouse zygotes.....	22
2.5 mESC culture prior to nucleofection .....	23
2.6 Nucleofection procedure .....	23
2.7 Immunohistochemistry and imaging .....	24
2.8 DNA extraction and genotyping.....	26
2.8.1 DNA extraction of embryos .....	26
2.8.2 DNA extraction of naïve mESCs .....	26
2.8.3 PCR reactions following DNA extraction followed by NGS.....	27
2.9 Shallow whole genome sequencing .....	28

2.10	Statistics.....	28
3.	Results.....	29
3.1	crRNA and primer design for each gene of interest.....	29
3.2	Optimization of CRISPR/Cas9 components in mouse ESCs.....	31
3.2.1	On-target efficiencies.....	31
3.2.2	Off-target effects.....	33
3.3	Immunofluorescent staining.....	35
3.4	Targeting <i>Gata3/Gata2</i> and <i>Tfap2a/Tfap2c</i> in mouse embryos via CRISPR/Cas9.....	37
3.5	Shallow whole genome sequencing to investigate (partial) chromosome loss.....	40
4.	Discussion.....	41
4.1	Optimization of the CRISPR/Cas9 components in mouse ESCs.....	41
4.2	Immunofluorescent staining of GATA2/3 at different embryonic stages.....	42
4.3	Targeting <i>Gata3</i> , <i>Gata2</i> , <i>Tfap2c</i> and <i>Tfap2a</i> with CRISPR/Cas9 in mouse embryos	43
4.3.1	On-target editing efficiency.....	43
4.3.2	Morphology of <i>Gata3</i> <i>-/-</i> , <i>Gata2</i> <i>-/-</i> , <i>Tfap2c</i> <i>-/-</i> and <i>Tfap2a</i> <i>-/-</i> mouse embryos.....	44
4.4	Shallow whole genome sequencing of WT, scrambled, <i>Gata2</i> and <i>Gata3</i> targeted naïve mESC.....	46
5.	Conclusion.....	47
6.	References.....	48
	Poster.....	57
	Supplementary.....	58
	Supplementary tables.....	58
	Supplementary figures.....	65

## Summary

**Background:** Subfertility can be overcome by application of assisted reproductive technologies (ART). However, implantation failure following embryo transfer occurs frequently, resulting in no pregnancy. Therefore, the function of important trophoctoderm (TE, lineage involved in implantation) markers *Gata2/3* and *Tfap2a/c* will be investigated in mouse embryos by targeted CRISPR/Cas9-mediated gene knock-out (KO).

**Materials and methods:** Upon assessment of the on-/ off-target effects of CRISPR/Cas9 in mESCs, CRISPR/Cas9 components were delivered to mouse zygotes using electroporation. Embryos were cultured until E4.5, followed by genotyping of the regions of interest. Expression patterns of GATA2/3 were determined using immunofluorescent (IF) staining. The potential occurrence of chromosome loss in mESCs was investigated using shallow whole genome sequencing (WGS).

**Results:** No significant off-target effects were observed in mESCs for all genes of interest, whereas *Gata2* *-/-*, *Gata3* *-/-*, *Tfap2a* *-/-* and *Tfap2c* *-/-* embryos could be created using the selected crRNAs. Embryos KO for our GOIs were able to form blastocysts, yet *Tfap2a* *-/-* and *Tfap2c* *-/-* blastocyst morphology was compromised. IF staining revealed earlier nuclear expression of GATA2 than GATA3 in WT mouse embryos. Shallow WGS of WT, *Gata2*, *Gata3* and scrambled targeted mESCs revealed chromosomal aberrations in all samples.

**Conclusion:** These results show the first indications of the role of *Tfap2a/c* in blastocyst formation, whereas *Gata2/3* show not to be crucial. Also, it was shown that nuclear GATA2 transcriptional activity potentially precedes GATA3 expression. The occurrence of CRISPR/Cas9 induced chromosomal aberrations should also be further investigated in both mESCs and embryos.

## Societal impact

Worldwide, 15% of couples suffer from subfertility, due to the contribution of several factors, such as maintaining a certain lifestyle, nutritional habits and the increasing age at which couples choose to conceive<sup>1,2</sup>. A possible solution to overcome this subfertility could be the application of assisted reproductive technologies (ART) such as *in vitro* fertilization (IVF) or intracytoplasmic sperm injection (ICSI). Although ART has changed the treatment of subfertility fundamentally and great progress has been made in this field, these techniques are not always successful as the biggest cause of ART failure remains the low implantation rate after embryo transfer<sup>3</sup>. To better understand unsuccessful implantation and placentation in this context, it is crucial to gain insight into the molecular mechanisms in the embryo's implantation and placentation processes. Many studies have already been done in this area using the mouse as a model system for the development of human embryos, yet several recent papers (reviewed in<sup>4</sup>) have revealed interspecies differences, putting a cautionary note on extrapolating findings done in mouse to human. Additionally, other models for human embryonic development such as blastoids, human embryonic stem cells (hESCs) and other mammal models (such as bovine and primate embryos) are considered suboptimal to study these molecular mechanisms during human embryonic development. This creates a necessity to also perform molecular studies directly in human embryos in the future.

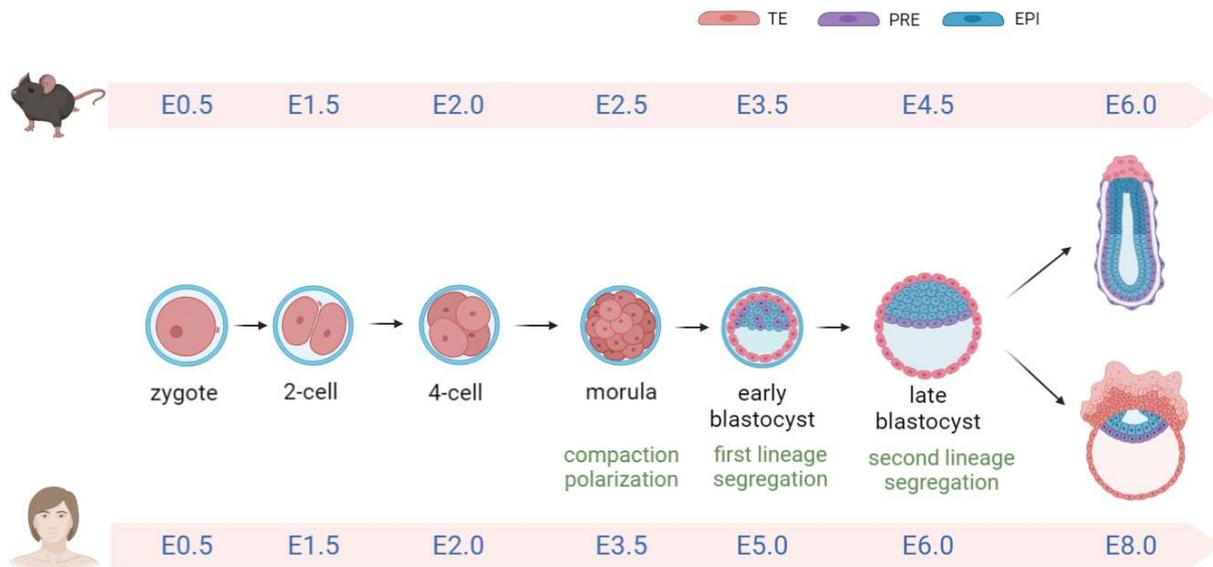
# 1. Introduction

## 1.1 Pre-implantation development

Following oocyte fertilization, numerous cellular events take place during mouse and human preimplantation development (**Figure 1**). Below, an overview is provided of the hallmark processes during mouse and human pre-implantation development.

Subsequent to fertilization, the totipotent zygote (embryonic day 0.5 (E0.5)) undergoes a series of mitotic (cleavage) divisions, resulting in an augmenting number of cells while maintaining a constant volume<sup>4</sup>. The first three mitotic divisions of the zygote sequentially give rise to a 2-cell (E1.5 for both mouse and human), 4-cell (E2.0 for both mouse and human embryos) and 8-cell embryo (E2.5 in mouse and E3.5 in human)<sup>5,6</sup>. At the 8-cell stage in mouse (E2.5) and between the 8 and 16-cell stage in human (E4.0), the processes of compaction and polarization occur, leading to the formation of a morula containing totipotent blastomeres<sup>5,6</sup>. Compaction is the first hallmark process taking place during the preimplantation development of the mouse and human embryo and is characterized by the metamorphosis of a loose accumulation of cells to a tightly packed cluster, due to an increased cell-cell contact between the blastomeres<sup>7</sup>. In the compacted mouse embryo, the outer blastomeres undergo polarization in a sequential manner by the acquisition of an apical domain, while the inner blastomeres remain apolar<sup>8</sup>. It is not clear whether a similar mechanism occurs in human embryos, as there are indications that compaction and polarization happen in a parallel manner<sup>9</sup>. Upon polarization in both mouse and human embryos, the polarized (outer) blastomeres will develop into trophectoderm cells (TE) while the apolar inner cells give rise to the inner cell mass (ICM) during the process which is referred to as the first lineage segregation, resulting in the formation of an early blastocyst (E3.5 in mouse and at E5.0 in human)<sup>5,6,8</sup>. The ICM, which contains pluripotent blastomeres, will give rise to the embryo proper, whereas the TE is responsible for the embryo's implantation and placentation<sup>5</sup>. Subsequently, during second lineage segregation, the ICM is divided into the epiblast (EPI, the embryonic lineage) and the primitive endoderm (PRE, the extra-embryonic lineage), contributing to the formation of a late blastocyst (E4.5 in mouse and at E6.0 in human)<sup>5,6</sup>. As opposed to the EPI, which gives rise to all fetal cells, the PRE primarily contributes to the extra-embryonic yolk sac<sup>6</sup>. Continuously, once the embryo has reached the late blastocyst stage, it hatches out of the zona pellucida (E4 in mouse and at E5-6 in human) to allow implantation and placentation of the embryo<sup>5,6</sup>.

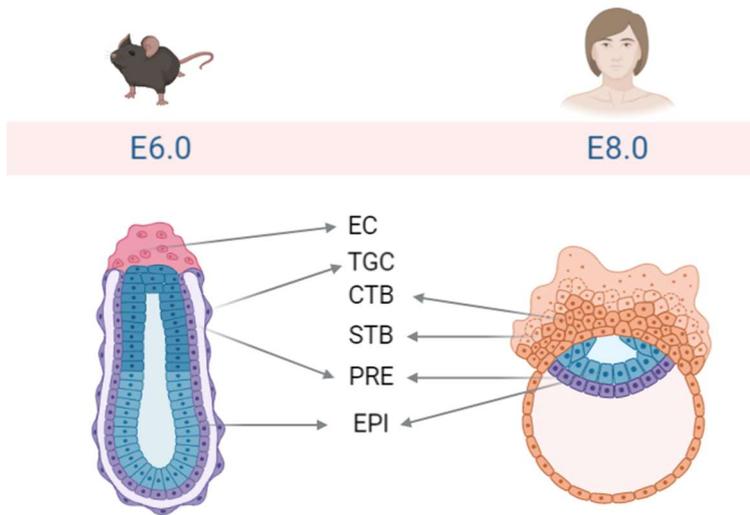
Given that we are interested in unravelling the molecular mechanisms involved in placentation and implantation of the embryo, we will research the molecular pathways underlying the specification, formation and maintenance of the TE (which is the lineage involved in implantation and placentation of the embryo).



**Figure 1: Pre-implantation embryonic development for mouse and human embryos.** black=stage of embryonic development, green=process occurring at that embryonic stage. Following fertilization, a zygote is formed. Multiple rounds of cleavages give rise to a morula, in which the processes of compaction and polarization occur. The inner cell mass (ICM) and trophectoderm (TE) are formed during the first lineage segregation, leading to the formation of an early blastocyst. Later, the ICM will segregate in the epiblast (EPI) and the primitive endoderm (PRE) during the second lineage segregation. At this stage, the late blastocyst hatches from the zona pellucida and is ready for implantation in the uterine wall.

## 1.2 Peri-implantation development

Once the blastocyst has hatched out of the zona pellucida, implantation and placentation of the embryo can occur<sup>5,6</sup>. After the blastocyst invades the uterine wall, complex interactions between differentiated TE cells and maternal cells in the decidua of the uterus will lead to the development of the placenta<sup>10</sup>. Placentation is crucial for the development of the embryo/fetus as the placenta facilitates the exchange of ions, metabolites and waste products between mother and fetus<sup>10</sup>. Both mouse and human post-implantation embryos are displayed in **Figure 2**. In mouse, the polar TE cells (TE cells in contact with the ICM), continue to multiply, resulting in the formation of the extra-embryonic ectoderm containing the cytotrophoblast cells (CTB) and the ectoplacental cone (EC) comprising syncytiotrophoblast cells (STB)<sup>11,12</sup>. The trophoblast giant cells (TGCs) (which facilitate attachment of the embryo to the uterine wall) are formed when TE cells on the opposite site of the ICM (mural TE) complete division<sup>11</sup>. In human, the CTB shell is formed by the mural TE, as it continues to grow<sup>13,14</sup>. The CTB cells are placental epithelial cells which can differentiate into extra villous trophoblast (EVT) cells, analogous to TGC in mouse, or STB cells<sup>13,14</sup>. The signal CTB cells receive, determines to which cell type they differentiate<sup>13</sup>. EVTs ensure attachment of the embryo to the uterine wall and remodel maternal spiral arterioles, whereas the STB is formed as a result of CTB fusion and secrete hormones required for pregnancy maintenance and act as a barrier for nutrient and gas exchange<sup>13</sup>.



**Figure 2: Peri-implantation mouse and human embryos.** In the mouse post-implantation embryo, the trophectoderm (TE) differentiates into the extra-embryonic ectoderm containing cytotrophoblast (CTB) cells and the ectoplacental cone (EC). Mural TE cells form the trophoblast giant cells (TGC). The human TE differentiates in the CTB and syncytiotrophoblast (STB), which will be responsible for the embryo's implantation and placentation.

### 1.3 Major genetic players during preimplantation development

On a molecular level, the processes of polarization, first and second lineage segregation and implantation are a result of complex interactions between several transcription factors (TFs), leading to differential gene expression. Although a lot of research has been conducted in the mouse in this context, the molecular mechanisms in human embryos remain mostly unraveled. Below, an overview will be provided of the key genetic players during these hallmark processes in mouse and human embryonic development.

#### 1.3.1 Polarization

In mouse embryos, polarization is the process in which the outer blastomeres become polarized by the acquisition of an apical domain, containing the protease activated receptor- atypical protein kinase C (Par-aPKC) protein complex and the ERM proteins (ezrin, radixin and moesin), enclosed by an actomyosin ring including F-actin<sup>8,15,16</sup>. Interestingly, formation of the apical domain in the outer blastomeres of the embryo differs from polarization in other cell types, as this can occur without any signals from an external source such as extracellular matrix or cell-cell adhesion<sup>8</sup>. For mouse, polarization happens at the 8-cell stage, following a large transcriptional wave<sup>8,17</sup>. The formation of the apical domain is directed by two distinct cooperative processes: actin-mediated cooperative recruitment and lateral mobility<sup>8</sup>. These two mechanisms shape the apical domain and act as antagonistic forces<sup>8</sup>. In this context, symmetry breaking and concentration of the apical proteins is enabled by actin-mediated cooperative recruitment, whereas their diffusion is allowed by the lateral mobility process<sup>8</sup>. These opposing processes result in a crescent-shaped apical domain in the outer blastomeres and need to be in balance, as unequal cooperation will result in a deformed apical domain<sup>8</sup>. Whereas elevated actin-mediated cooperative recruitment forces would result in small and numerous apical domains, excessive lateral mobility forces would produce an invariable distribution of apical proteins, inhibiting symmetry breaking<sup>8</sup>. It has been shown that TFs AP2γ and TEA domain transcription factor 4 (TEAD4) (encoded by the murine *Tfap2c* and *Tead4* genes respectively) are crucial for the actin-mediated cooperative recruitment in the outer blastomeres of mouse embryos<sup>8</sup>. Both TFs support this process by controlling actin remodeling to direct the growth of apical protein clusters by regulating the expression of genes such as *Rock*, *Arp2/3* and *Pard6b*<sup>8</sup>. Actin regulators encoded by these genes (ROCK, ARP2/3 and PARD6B) direct the membrane recruitment of apical proteins<sup>8</sup>.

Alongside, Ras homology family member A (RhoA) is involved in lateral mobility and permits the formation of the apical domain <sup>8</sup>.

Until recently, the mechanism and timing of polarization of the outer blastomeres of the human embryo remained unclear, as well as whether this process was conserved across species <sup>9</sup>. Lately, it has been reported that in human embryos, polarization is driven by a similar process as in mouse embryos as polarization in the outer blastomeres of human embryos also occurs by a two-step mechanism <sup>9</sup>. First, apical enrichment of F-actin takes place (concomitantly with compaction), followed by the apical accumulation of the Par-aPKC complex <sup>9</sup>. Both processes lead to the formation of an apical domain in the outer blastomeres of the human embryo at the 8-16 cell stage <sup>9</sup>. These results suggest a conserved mechanism behind polarization of the outer blastomeres in mice and human embryos <sup>9</sup>. However, in mouse embryos, compaction and polarization seem to occur sequentially, while recent studies underline the possibility that there is certain overlap between compaction and polarization in human embryos, as there is accumulation of F-actin at the apical side of the outer blastomeres concomitant with compaction <sup>8,9</sup>. This would imply that compaction and polarization happen in a parallel manner rather than sequentially, as seen in mouse <sup>8,9</sup>.

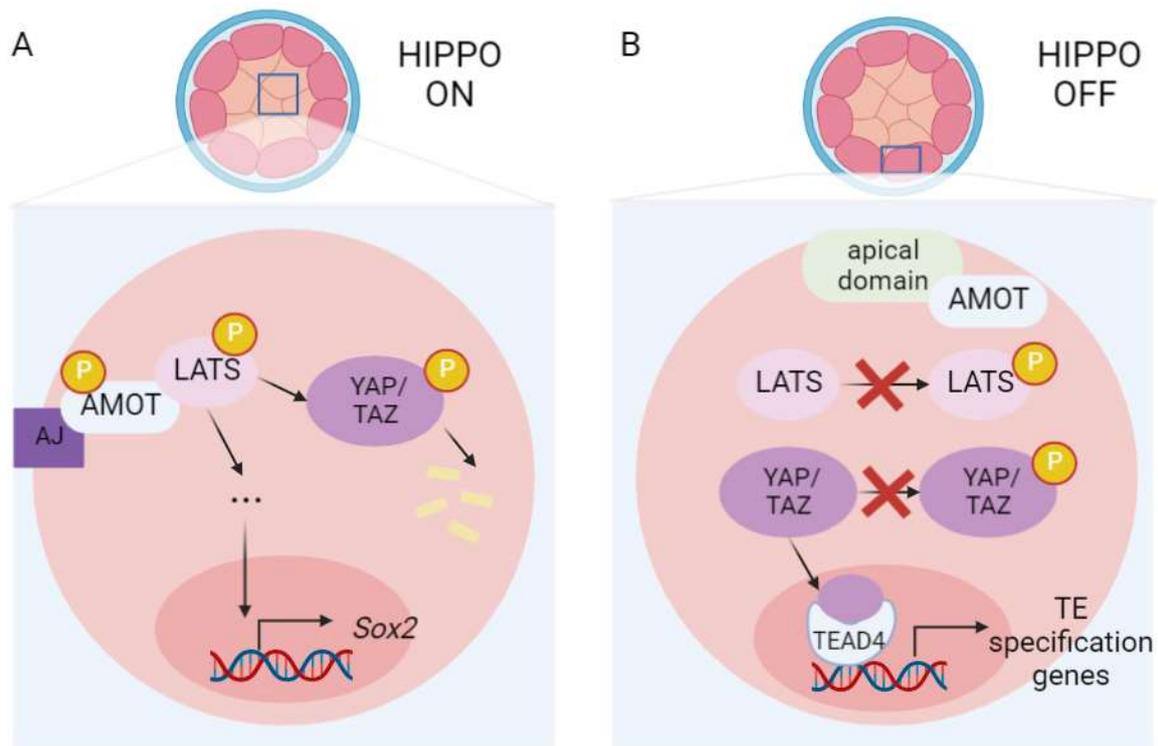
### 1.3.2 First lineage segregation

In the mouse embryo, once the apical domain is formed in the outer blastomeres, the Par-aPKC protein complex will restrict the protein angiomin 1 (AMOT) to the apical domain <sup>18</sup>. Distribution of AMOT is different in outer and inner blastomeres as in the inner blastomeres, AMOT is localized at the adherens junctions (AJs) of the plasma membrane whereas in the outer blastomeres, AMOT is restricted to the apical domain (**Figure 3**) <sup>19</sup>. When AMOT is present near the AJs of the plasma membrane, it is in close proximity of the Large tumor suppressor kinase (LATS), which phosphorylates Serine 176 of the N-terminal part of AMOT (**Figure 3A**)<sup>18,19</sup>. This post-translational modification (PTM) will stabilize the interaction between AMOT and LATS, resulting in the activation of LATS, allowing the Hippo pathway to be switched on <sup>18,19</sup>. Once activated, LATS will phosphorylate the Yes associated protein/transcriptional co-activator with PDZ binding motif complex (YAP/TAZ complex), resulting in its degradation <sup>19</sup>. Therefore, the YAP/TAZ complex will not be capable to translocate to the nucleus <sup>19</sup>. Moreover, when the Hippo pathway is switched on in this context, LATS remains active and will facilitate the expression of the *Sox2* gene – which is considered the first marker of ICM pluripotency <sup>20</sup>. The reason behind the early expression of *Sox2* remains unclear, as *Sox2* only plays an autonomous role in cell fate in the late mouse blastocyst stage <sup>20</sup>. A potential explanation could be that *Sox2* is genetically redundant with other pluripotency factors (such as *Oct4* and/or *Nanog*) during these early stages in development <sup>20</sup>. During first lineage segregation, TFs such as SRY-box transcription factor 2 (SOX2), Nanog Homeobox (NANOG) and octamer-binding transcription factor 4 (OCT4) become upregulated in the ICM <sup>21-23</sup>. They promote the pluripotency and inhibit differentiation of the ICM cells as OCT4 is crucial to prevent diversion of the ICM towards TE lineage whereas NANOG is also essential for maintaining pluripotency <sup>22,24,25</sup>. Moreover, NANOG also prevents the ICM cells from differentiating towards PRE during second lineage segregation <sup>22,25</sup>.

In the outer blastomeres, AMOT is restricted to the apical domain by the Par-aPKC protein complex, and thus will not be phosphorylated by LATS (**Figure 3B**)<sup>18,19</sup>. In return, LATS will not be activated and the Hippo pathway is switched off <sup>19</sup>. As a consequence, YAP/TAZ will not be phosphorylated and instead, translocates to the nucleus<sup>19</sup>. In the nucleus, the YAP/TAZ complex will act as a transcriptional co-activator for the TF TEAD4 <sup>19,26</sup>. This

ultimately leads to the transcription of TE specification genes such as *Cdx2*, *Gata3*, *Eomes* and *Tfap2a/c*<sup>19</sup>, which are crucial for implantation and placentation of the embryo<sup>24</sup>.

Additionally, it was reported that RHO and ROCK negatively regulate Hippo pathway signaling in the outer blastomeres of the mouse embryo<sup>8,24,27,28</sup>. First of all, RHO and ROCK can regulate the subcellular localization of LATS activators (such as AMOT) and thereby inhibit activation of LATS<sup>24,27,28</sup>. Also, RHO can directly regulate the localization of the apical domain, leading to the inactivation of the Hippo signaling<sup>8</sup>. Furthermore, it has been shown that AP2 $\gamma$  positively regulates ROCK expression, which negatively controls the activity of the Hippo signaling and thus promotes TE commitment<sup>8,24</sup>.



**Figure 3: Molecular mechanisms involved in polarization in inner and outer blastomeres in mouse embryos.** (A) **Inner blastomere.** As AMOT is positioned at the adherens junctions (AJs), LATS can phosphorylate AMOT, leading to stabilization of the interaction between AMOT and LATS and the activation of LATS kinase. The YAP/TAZ complex will be phosphorylated and subsequently degraded. Consequently, Sox2 – which is the earliest inner cell mass (ICM) pluripotency marker - will be expressed, leading to the differentiation of these cells into the ICM. (B) **Outer blastomere.** AMOT is restricted to the apical domain, and therefore, LATS cannot phosphorylate AMOT. Thus, LATS kinase will not get activated and will not be able to phosphorylate the YAP/TAZ complex. Now, the YAP/TAZ complex is able to translocate to the nucleus to act as a transcriptional co-activator of transcription factor (TF) TEAD4. This leads to the transcription of trophectoderm (TE) specification genes and TE commitment.

Moreover, an early role for bone morphogenic protein 4 (BMP4) signaling in the TE specification of the mouse embryo is suggested<sup>29,30</sup>. At the 16-cell stage, TE cells in mouse embryos exhibit enrichment of the bone morphogenic protein 2 receptor (BMP2R) and establish phosphorylated Suppressor of Mothers Against Decapentaplegic 1 (SMAD1, downstream marker of activated BMP4 signaling)<sup>29</sup>. Meanwhile, the ICM shows expression of BMP4/7<sup>29</sup>. These phenomena indicate paracrine signaling from the ICM to the TE in which BMP4 ligand will be secreted from the ICM and will bind to the BMP2R in the TE cells<sup>29</sup>. Besides, treatment of mouse embryos with BMP4 will result in morula arrest<sup>30</sup>. When this was analyzed more in detail, it was reported that BMP4 induced translocation of TEAD4 to

the nucleus, causing transcription of TE specification genes such as *Cdx2*, *Gata3* and *Tfap2c*<sup>30</sup>. This was also induced in the inner blastomeres (future ICM cells), leading to aforementioned morula arrest<sup>30</sup>. In conclusion, these results suggest a role of BMP4 in the TE specification in the mouse preimplantation embryo<sup>29,30</sup>. Furthermore, a study has been conducted by De Paepe *et al.*, in which human embryos cultured *in vitro* were supplemented with BMP4<sup>31</sup>. Treatment of human preimplantation embryos with BMP4 resulted in impaired blastocyst development, yet this was not due to interference with TE formation, as expression levels of CDX2 and GATA3 remained similar to non-treated control human embryos<sup>31</sup>. Administration of BMP4 rather results in apoptosis in human preimplantation embryos<sup>31</sup>. However, De Paepe *et al.* also mention that it could be possible that another BMP factor is involved in human TE formation, rather than BMP4<sup>31</sup>. On the contrary, BMP4 may play a role in processes that occur before the first lineage segregation, though this has not been investigated to date.

It is still unknown if the first lineage segregation in human preimplantation results from a similar mechanism<sup>9</sup>. Recent research of Zhu *et al.* has revealed that while polarization strengthens TE specification of the outer blastomeres, TE factors like CDX2 and GATA3 are expressed independently from the process that follows polarization in human embryos<sup>9</sup>. These findings imply that Hippo signaling might contribute to the first lineage segregation in human embryos, yet it is also possible that other mechanisms and/or pathways are essential to this process, as polarization can enhance TE specification, but it is not necessary for this process in human embryos<sup>9</sup>.

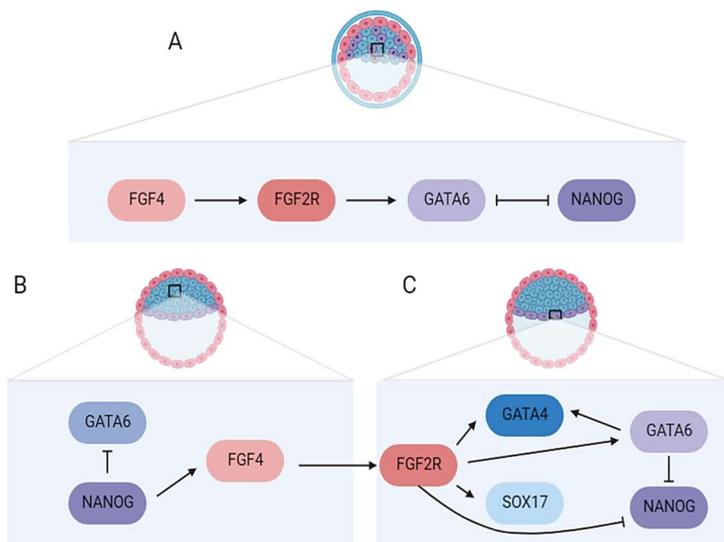
### 1.3.3 Second lineage segregation

In mouse, throughout the whole ICM, there is a combination of cells destined to be specified to the EPI or PRE lineage<sup>4</sup>. In mouse, the heterogeneity of PRE and EPI precursor cells in the ICM of the early blastocyst is established through several rounds of cell internalization<sup>4</sup>. The EPI precursors are formed from early internalized cells that express higher levels of FGF4 ligands, whereas PRE precursors originate from later internalized cells that exhibit higher levels of FGFR2<sup>4</sup>.

Before the second lineage segregation, throughout the whole ICM, fibroblast growth factor 4 (FGF4) and fibroblast growth factor receptor 2 (FGF2R) will be expressed in all cells (**Figure 4A**)<sup>32</sup>. During differentiation of the ICM in EPI cells, *Nanog* expression will be upregulated, resulting in the upregulation of FGF4 in the EPI precursor cells, whereas FGF2R expression will be lost<sup>32</sup> (**Figure 4B**). Therefore, no FGF4 signaling can be established in the future EPI cells, which is necessary for PRE specification<sup>32</sup>. Altogether, GATA6 expression (which is a PRE specification marker) is completely lost in the future EPI cells, as a result of *Nanog* upregulation and FGFR2 downregulation<sup>32</sup>. This mechanism results in the formation of EPI cells. In addition, together with NANOG, OCT4 and SOX2 will further specify and maintain the EPI<sup>32</sup>. Furthermore, in the future PRE cells, the opposite mechanism occurs (**Figure 4C**)<sup>32</sup>. Due to *Gata6* upregulation, *Nanog* expression will be lost, resulting in the upregulation of FGFR2<sup>32</sup>. FGF4, originating from the EPI, will bind to FGFR2, leading to the further enhancement of GATA6 expression<sup>32</sup>. In addition, FGFR2, and an autoregulation of GATA6 ensure the repression of *Nanog* expression in future PRE cells, and upregulation of PRE specification markers such as SOX17 and GATA4<sup>32</sup>. Additionally, a regulatory feedback loop between FGFR2, GATA6, SOX17 and GATA4 will be necessary to maintain PRE identity<sup>32</sup>.

In human, it seems that EPI, PRE and TE lineages appear to form distinct molecular profiles only after the blastocyst is formed<sup>4,33</sup>. While EPI and PRE lineages are clearly distinguished, a mature PRE layer does not become visible until the late blastocyst stage, at the beginning

of implantation <sup>4,34</sup>. It remains unclear whether the mechanism behind second lineage segregation in human remains the same as in mouse embryos <sup>4</sup>. It has been suggested that other signaling pathways are underlying the second lineage segregation in human embryos, as inhibition of FGF signaling does not result in preventing the formation of the PRE <sup>4,35</sup>.



**Figure 4: Molecular mechanisms during second lineage segregation in mouse.** (A) In the inner cell mass (ICM), there is a heterogeneity of cells, destined to become epiblast (EPI) or primitive endoderm (PRE). Gata6 expression results in inhibition of Nanog and vice versa. (B) In the EPI, NANOG will be expressed together with FGF4, but there is loss of FGF2R expression. FGF2R is necessary for the GATA6 expression and because NANOG will also inhibit GATA6 expression, there is loss of GATA6. (C) In the PRE, FGF2R will ensure transcription of GATA6 which leads to the inhibition of NANOG expression. Also, PRE specification genes such as Sox17 and Gata4 will be transcribed. There is a positive feedback loop between GATA4, SOX17 and GATA6 to ensure PRE maintenance.

#### 1.3.4 Implantation

In the mouse embryo, the mural TE will regulate adhesion of the embryo to the uterine wall <sup>4</sup>. Due to downregulation of *Cdx2* expression in the mural TE prior to implantation, a physical barrier is established between the polar and mural TE <sup>4,36</sup>. As opposed to the polar TE, cells of the mural TE undergo an epithelial to mesenchymal transition, as a consequence of multipotency loss <sup>4,37</sup>. Within the TGCs,  $\alpha\beta3$  receptors will mitigate the first contact with the uterine wall, resulting in activation of calcium signaling <sup>4</sup>. This ensures further integrin recruitment to enhance implantation of the mouse embryo <sup>4,38</sup>.

In human, the implantation process is controlled by the polar TE <sup>4,39,40</sup>. Upon the initial contact between the uterine wall and the embryo, the polar TE multiplies and differentiates into the CTB and STB, which results in the implantation of the blastocyst <sup>4,41,42</sup>. Besides, it has been discovered that T-box transcription factor 3 (TBX3) is a key player in controlling trophoblast differentiation to CTB and STB <sup>4,43</sup>.

#### 1.4 Interspecies differences

As illustrated above, extensive research has been performed regarding molecular mechanisms involved in embryo implantation and placentation, using mouse as a model for human embryonic development. However, a number of recent studies highlight important interspecies differences between mouse and human pre- and post-implantation embryonic development (reviewed in <sup>4</sup>). These findings put a cautionary note on extrapolating the findings from mouse embryos directly to human. Below, a summary will be provided of the main interspecies differences between mouse and human pre-implantation embryonic development.

The first important hallmark processes in both mouse and human embryonic development are compaction and polarization <sup>8</sup>. As mentioned above, once compaction occurs in the pre-

implantation embryo, the outer blastomeres acquire an apical domain through the process of polarization<sup>8</sup>. An important difference in timing of compaction and polarization exists between mouse and human embryos<sup>4</sup>. In mouse, polarization and compaction of all blastomeres occurs by the end of the 8-cell stage while in human embryos, this only happens later between the 8- and 16-cell stage<sup>4,44,45</sup>. However, since certain human embryos finish compaction and polarization by the end of the 8-cell stage, whereas other embryos don't finish these processes until after the 8-cell stage, the timing of polarization and compaction in human embryos appears to be less uniform<sup>9</sup>. Furthermore, in mouse embryos, generation of the inner cells occurs only after all outer blastomeres become polarized, while in human embryos, inner cells can be detected in embryos that are not fully polarized<sup>7-9</sup>. This could point out that polarization and compaction occur in a parallel manner in the human embryo, opposing to the polarization and compaction processes in mouse, which happen sequentially<sup>7-9</sup>. These findings might suggest that the first lineage decision may differ in a mechanical manner between mouse and human preimplantation embryos<sup>9</sup>.

In mouse, following polarization, differential activation of the Hippo pathway in the outer and inner blastomeres leads to the specification of the TE and the ICM fates during first lineage segregation<sup>18,19</sup>. Yet, it is still unknown if a comparable mechanism is responsible for the first lineage segregation in human embryos<sup>9</sup>. It has recently been reported that PLC $\zeta$  signaling acts upstream of the polarization of the outer blastomeres in human embryos<sup>9</sup>. Researchers also discovered that the expression of *GATA3* could occur independently of polarization, whereas downregulating key enzymes in PLC $\zeta$  signaling reduced *GATA3* expression by lowering the proportion of polarized cells<sup>9</sup>. These findings suggest that apical domain formation reinforces the TE fate<sup>9</sup>. Yet, the human polarization process seems to differ from the mechanism in the mouse embryo, where switching off of the Hippo pathway due to formation of the apical domain in the outer blastomeres exclusively induces the expression of *GATA3* via the YAP/TAZ complex and the TF TEAD4<sup>19,26</sup>. Moreover, these results could imply that in human embryos, Hippo signaling is involved in the first lineage segregation, nonetheless concomitant with other signaling pathways and/or mechanisms<sup>9</sup>. Altogether, these findings possibly demonstrate a different mechanism underlying polarization in mouse and human preimplantation embryos<sup>9</sup>. Moreover, it was reported by Stamatiadis *et al.* that TEAD4 also induces TE specification in the human embryo by upregulation of *CDX2*<sup>46</sup>. However, it must be noted that, as opposed to mouse embryos, TE specification occurred in a *GATA3* independent way<sup>46</sup>. It was therefore proposed that in order to specify the TE in human embryos, the *CDX2/TEAD4* mechanism works in parallel with *GATA3*<sup>46</sup>.

When further investigating expression of TE and ICM markers during first lineage segregation in both mouse and human embryos, it has been shown that in human, *OCT4* is critical for the expression of the TE specification marker *CDX2*<sup>47,48</sup>. This is in contrast to mouse TE specification since *Cdx2* is expressed following the switching off of the Hippo pathway in the outer blastomeres<sup>19,48</sup>. It was also confirmed that *OCT4* expression is required earlier in human embryonic development as developmental arrest occurred at an earlier embryonic stage in *OCT4* *-/-* human embryos compared to *Oct4* *-/-* mouse embryos<sup>48</sup>. Furthermore, the human TE is specified independently of *CDX2* expression, which does not increase until after blastocyst formation<sup>4,34</sup>. These results support the findings that both *CDX2/TEAD4* and *GATA3* are involved in TE formation in a parallel manner, as *CDX2* expression is observed not to be the only mechanism underlying TE specification in the human embryo<sup>46</sup>. In addition, these findings might imply that even though Hippo signaling seems to be conserved in human embryos, the first lineage segregation in the human embryo is potentially regulated by an additional molecular mechanism which is not necessary for first lineage segregation in the mouse embryo<sup>4,9</sup>.

When comparing second lineage segregation in mouse and human embryos, it was reported that in mouse, FGF4 ligand secretion (originating from the EPI) activates FGF signaling in cells expressing FGFR2 (PRE precursor cells), resulting in the expression of PRE markers such as GATA6<sup>4,49</sup>. Because of this phenomenon, the GATA6-positive and NANOG-positive cells will respectively form the PRE and the EPI cells in the mouse embryo<sup>4,49</sup>. In human embryos, it is possible that another mechanism is controlling the second lineage segregation since PRE (hypoblast) specification is not blocked following inhibition of FGF signaling<sup>4,35</sup>. Also, it was reported in human embryos that there is a downregulation of GATA6 in a subset of the SOX17-positive cells during second lineage segregation, in contrast to mouse embryos<sup>34</sup>. Both observations point to a distinct molecular mechanism controlling human PRE specification<sup>32,34</sup>. Moreover, in human embryos, specification of EPI, TE and PRE on a molecular level only occur after blastocyst formation<sup>4,33</sup>.

Furthermore, single-cell RNA sequencing of the three lineages in mouse and human embryos, performed by Blakeley *et al.*, has revealed significant differences in gene expression in human preimplantation development in comparison to mouse<sup>50</sup>. First, several interspecies differences regarding gene expression in the TE were revealed as in mouse, key TFs in TE specification (*Elf5*, *Eomes Id2*) are enriched in the TE while in human, *ELF5*, *ID2* and *EOMES* expression is absent in the TE<sup>50</sup>. Additionally, expression of several genes expressed in the human TE (*CLDN10*, *TRIML1* and *PLAC8*) were absent in mouse TE, suggesting distinct mechanisms regarding TE formation in human embryos<sup>50</sup>. In contrast to human embryos, *Sox2* is also co-expressed with a number of TE specification genes in mouse, whereas this is not reported in human<sup>50</sup>. Yet, the finding which was of most interest for our study, is that differential expression of *Tfap2c/TFAP2C* (a key regulator in TE specification in the mouse) between mouse and human embryos is also detected, yet as this protein is the focus of this research project, this will be explained later on more in detail (see section 1.5.1)<sup>50</sup>. Moreover, differential gene expression between mouse and human ICM was also served<sup>50</sup>. When looking at the genes that are co-expressed with *Oct4/OCT4* in both mouse and human, it was discovered that mouse and human embryos express different genes in conjunction with *Oct4/OCT4*<sup>50</sup>. Besides, when *Sox2/SOX2* expression was investigated in both mouse and human embryos, it was found that *Sox2* expression is upregulated at the blastocyst stage in the mouse embryo while *SOX2* is abundantly expressed at the 4-cell to blastocyst stage in the human embryo<sup>50</sup>. Interestingly, timing of *Nanog/NANOG* expression in mouse and human embryos also differs<sup>50</sup>. These phenomena point to the fundamental variations in gene expression during first and second lineage segregation in mouse and human embryos<sup>50</sup>.

In addition, the mechanism responsible for embryo implantation also differs between mouse and human<sup>4</sup>. In the TE of both mouse and human embryos, a distinction can be made between the mural TE and the polar TE<sup>4</sup>. In mouse, the process of adhesion to the uterine wall and implantation is regulated by the mural TE and occurs at E4.5-E6<sup>4</sup>. In contrast, in human embryos, the adhesion and implantation of the embryo to the uterine wall is regulated by the polar TE and occurs at E7 and E8<sup>4,51</sup>. As opposed to human embryos, the mouse embryo's polar TE preserves its multipotent nature as a result of FGF signaling from the EPI, and is not involved in implantation and placentation of the mouse embryo<sup>4,52</sup>.

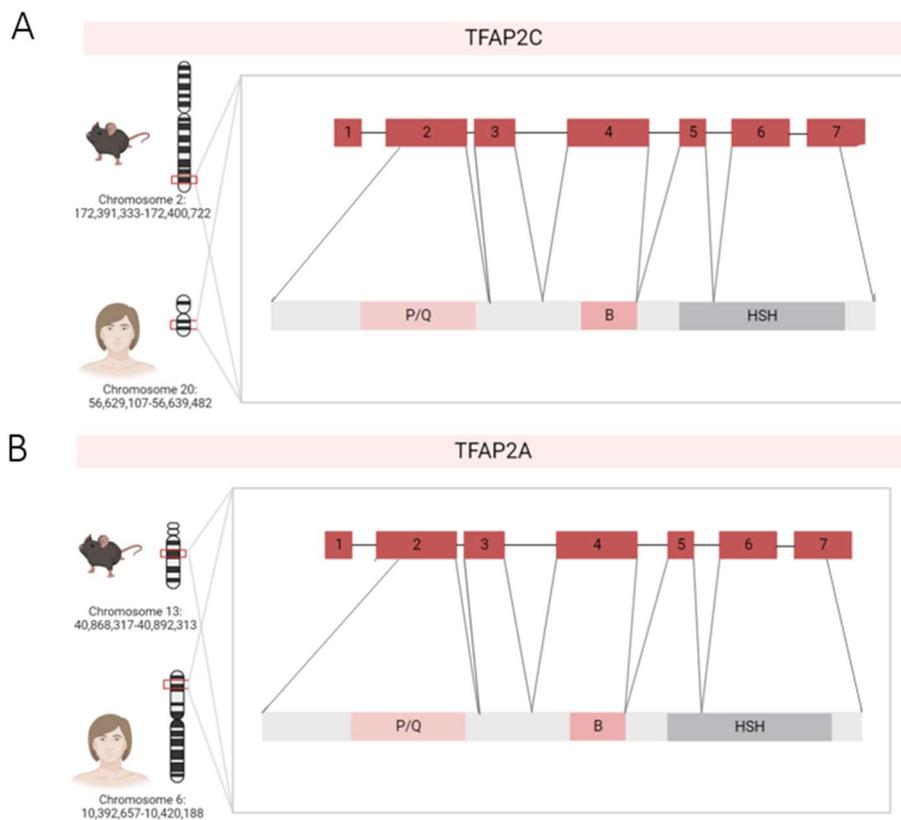
These findings demonstrate the difficulty of extrapolating results from research in mouse embryos to human embryos. The aforementioned results also highlight the need for research into molecular mechanisms directly in human embryos. To maximize comparison between the two models, molecular mechanisms involved in the formation of the preimplantation mouse and human embryo must be researched in the same study, as a lot of technical variation exists.

## 1.5 Genes of interest

Based on single-cell RNA-sequencing datasets from both Blakeley *et al.*<sup>50</sup> and Petropoulos *et al.*<sup>33</sup>, we identified several genes coding for important TE markers which are highly conserved in both mouse and human embryos, and of which the function will be investigated. *Gata3*, *Tfap2c* and the highly similar TFs *Gata2* and *Tfap2a* will be investigated in mouse (and ideally in human embryos on the long term) to get a better insight in the events prior to implantation and placentation of the embryo. On the long term, results from this study could provide us a clearer understanding of implantation failure of the embryo following embryo transfer after ART. Additionally, increased support of *in vitro* cultured embryos could also be reached. As displayed above, due to significant interspecies differences, it is not always possible to extrapolate results from mouse embryos to human. As a result, it is desired to assess if interspecies differences occur in the first lineage segregation between mouse and human embryos, on the long term.

### 1.5.1 *Tfap2a/TFAP2A*, *Tfap2c/TFAP2C*

In both mouse and human, *Tfap2a/TFAP2A* and *Tfap2c/TFAP2C* code for the TFs AP2 $\alpha$  and AP2 $\gamma$  respectively, both belonging to the activating enhancer binding protein 2 (AP2) family<sup>53</sup>. Upon activation, TFs of the AP2 family can form homo- or heterodimers which actively regulate gene transcription after binding guanine/cytosine (G/C) rich regions in the DNA of the target gene<sup>53</sup>. These G/C rich sequences are found in various cellular enhancers<sup>53</sup>. All members of the AP2 family have a highly conserved protein structure in both mouse and human (**Figure 5**)<sup>53</sup>. Once these TFs are activated, the C-terminal helix-span-helix (HSH) motif regulates homo- or heterodimerization, while the central basic region (B) (in cooperation with the HSH motif) regulates the binding of these homo- or heterodimers to the target DNA after dimerization<sup>53</sup>. The N-terminal region, abundant in proline and glutamine residues, contains the transactivation proline/glutamine (P/Q) rich domain<sup>53</sup>. However, when investigating the protein sequence of the TFAP2A/C proteins in both mouse and human, no typical nuclear localization signal (NLS) was found<sup>54</sup>. Yet, a certain region in the DNA binding domain (AA 254-285) is predicted to fulfill the function of nuclear localization<sup>54</sup>. Another possibility could be that AP2 $\alpha/\gamma$  has the ability to bind to other proteins containing an NLS, and therefore are able to translocate to the nucleus. However, the latter has not been investigated.



**Figure 5: Protein structure of TFAP2A and TFAP2C, alongside coding exons for each domain. (A) Protein structure of TFAP2C (B) Protein structure of TFAP2A. In both mouse and human, exon 2 codes for the N-terminal (proline/glutamine) P/Q rich domain, while exon 4 codes for the (basic) B-domain. Additionally, exons 5, 6 and 7 code for the (helix-span-helix) HSH domain, responsible for the homo-or heterodimerization of the transcription factors (TFs).**

In mouse, the *Tfap2a* gene (**Figure 5B**) is located on chromosome 13 (40 868 778- 40 883 919), contains seven exons and nine transcripts are known (**Supplementary Table 1**)<sup>109</sup>. Generally, exon 2 will code for the P/Q rich transactivation domain while exon 4 codes for the central basic domain<sup>55,56</sup>. The N-terminal HSH motif, which is responsible for homo-or heterodimerization, is encoded by exons 5,6 and 7<sup>57</sup>. Aside from that, the *TFAP2A* gene (**Figure 5B**) in humans has seven exons, is found on chromosome 6 (10 392 657-10 420 188) and has been linked to four different transcripts (**Supplementary table 2**)<sup>110</sup>. In the human *TFAP2A* gene (**Figure 5B**), exon coding for the same domains correspond to the murine *Tfap2a* gene<sup>57</sup>.

The *Tfap2c* gene in mice (**Figure 5A**) has seven exons, is located on chromosome 2 (position 172 391 513 – 172 400 542) and has been associated with six different transcripts (**Supplementary Table 3**)<sup>111</sup>. Comparing this gene to the *Tfap2a* mouse gene (**Figure 5B**), the exons encoding the various domains are generally conserved<sup>58</sup>. Additionally, the human *TFAP2C* gene (**Figure 5A**) has seven exons, is located on chromosome 20 (55 629 107-56 639 482) and has two known transcripts (**Supplementary Table 4**)<sup>112</sup>. Moreover, when comparing the mouse and human *Tfap2c/TFAP2C* gene, it can be observed that exon coding is conserved in human<sup>58</sup>.

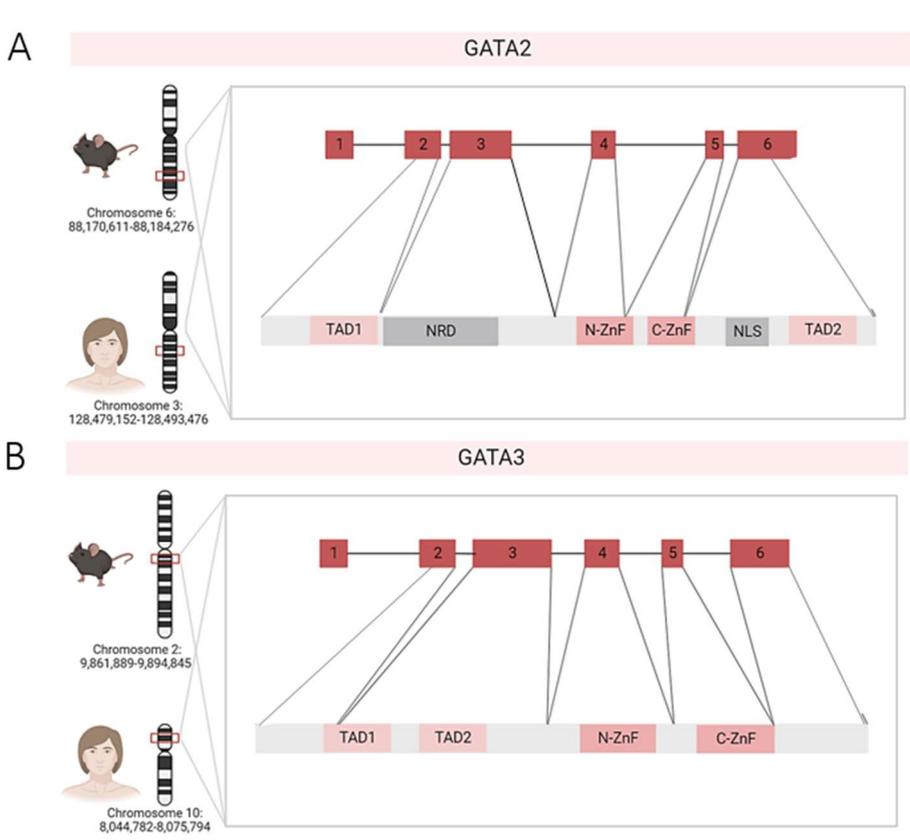
Both *Tfap2a/TFAP2A* and *Tfap2c/TFAP2C* show many functions throughout embryonic development in both mouse and human embryos<sup>8,19,59-62</sup>. As mentioned above, in mouse, *Tfap2c* expression is upregulated concomitant with *Tead4* to ensure formation of an apical domain in the outer blastomeres<sup>8</sup>. AP2γ and TEAD4 will regulate actin mediated cooperative recruitment, necessary for the formation of the apical domain<sup>8</sup>. Furthermore, in mouse,

*Tfap2a* and *Tfap2c* are expressed once the apical domain formation leads to the switching off of the Hippo pathway in the outer blastomeres, indicating that these genes are crucial for TE specification and differentiation in mouse embryos<sup>19</sup>. Corresponding with mouse embryonic development, it was discovered that *TFAP2C* is part of a core regulatory network of TFs, essential for the differentiation and maintenance of the TE of human embryos, which also controls the growth of the human placenta<sup>59,60</sup>. Moreover, in *Tfap2a/Tfap2c* double KO mouse embryos, lethality is earlier than for the *Tfap2c* single KO embryos, suggesting that both genes have redundant functions<sup>60,61</sup>. In addition, at the post-implantation level, the support of lineage identity, proliferation and migration of the CTB in human embryos depends on the expression of *TFAP2C*<sup>60,62</sup>. Finally, the post-mitotic STB exhibits *TFAP2A* expression, which controls the expression of placental hormones<sup>60,62</sup>.

*Tfap2c* has been identified as a key regulator of TE specification during first lineage segregation in the mouse<sup>8</sup>. However, single-cell RNA sequencing of the three lineages in both mouse and human embryos has revealed different expression patterns of *Tfap2c/TFAP2C* during mouse and human embryonic development<sup>50</sup>. In mouse, AP2 $\gamma$  is first detected at the zygote stage and expression levels remain high throughout embryonic development<sup>50</sup>. In human, AP2 $\gamma$  is only detected at the 8-cell stage<sup>50</sup>. Furthermore, *Tfap2c* is enriched in mouse TE as this gene was identified to be a crucial regulator of TE specification<sup>50</sup>. This is in contrast to human embryos, as *TFAP2C* is expressed in similar levels in both TE and EPI<sup>50</sup>. When investigating mouse blastocysts, AP2 $\gamma$  was only found in CDX2+ TE cells and was undetectable in NANOG+ EPI cells<sup>50</sup>. Yet, in contrast to mouse, AP2 $\gamma$  was present in both CDX2+ TE cells and NANOG+ EPI cells in human blastocysts<sup>50</sup>. This further suggests that there are fundamental differences in the molecular mechanisms underlying first lineage segregation during mouse and human preimplantation development. As interspecies differences are strongly supported by the abovementioned studies, this also stresses the necessity to investigate the function of both genes in mouse and human embryos.

### 1.5.2 *Gata2/GATA2, Gata3/GATA3*

In both mouse and human, GATA2 and GATA3 are both members of the GATA family of transcriptional regulatory proteins which will bind a GATA-motif in the target DNA in order to regulate gene transcription<sup>63</sup>. Upon activation, GATA2 and GATA3 form homo- or heterodimers, which is important for the combinatorial and synergistic transcription regulation<sup>64</sup>. Both TFs contain two highly conserved zinc finger (ZnF) domains, which are responsible for DNA binding and which are involved in interactions with cofactors and other transcriptional factors, in addition to other protein-protein interactions<sup>63</sup>. Depending on whether the ZnF domains are located closer to the N- or C-terminal end of the protein, the two ZnFs are referred to as the N- and C-terminal ZnF domains. In addition, GATA2 also contains a nuclear localization signal (NLS), a negative regulatory domain (NRD) and both an N- and C-terminal transactivation domain (TAD) in both mouse and human (**Figure 6A**)<sup>63,65</sup>. Besides, the *Gata3/GATA3* gene codes for a protein that, in addition to the highly conserved N- and C-terminal ZnF domains, also contains two TAD domains<sup>66</sup> (**Figure 6B**).



**Figure 6: Protein structure of GATA2 and GATA3 in both mouse and human. (A) Protein structure of GATA2, alongside exon coding for different domains. The GATA2 protein maintains the same domain structure in mice and humans. The protein consists of an N-terminal transactivation domain (TAD), a negative regulatory domain (NRD), an N- and C-terminal zinc finger domain (ZnF) a nuclear localization signal (NLS) and a C-terminal TAD. (B) Protein structure of GATA3, alongside exon coding for different domains. The GATA3 protein also maintains the same domain structure in mouse and human and contains two TAD (TAD) and two ZnF domains.**

In mouse, the *Gata2* gene is located on chromosome 6 (position 88 170 873-88 184 014), contains six exons, and five different transcripts are known for this gene in the mouse (**Supplementary table 5**)<sup>113</sup>. **Figure 6A** displays exon coding for the specific domains, alongside its protein structure in both mouse and human. Besides, the human *GATA2* gene is located on chromosome 3 (position 128 479 427 – 128 493 201) and contains six exons<sup>109</sup>. Furthermore, 13 different transcripts are known for *GATA2* (**Supplementary table 6**)<sup>114</sup>.

Five different transcripts have been associated with the mouse *Gata3* gene, which is positioned at chromosome 2 (9 861 889-9 894 845) and contains six exons (**Supplementary table 7**)<sup>115</sup>. **Figure 6B** displays the protein structure and exons that code for the various domains in the *GATA3* protein in both mouse and human. In addition, the human *GATA3* gene has six exons and is located on chromosome 10 (position 8 045 378 – 8 075 198)<sup>116</sup>. Also, five different transcripts are linked to the *GATA3* gene in human (**Supplementary table 8**)<sup>116</sup>.

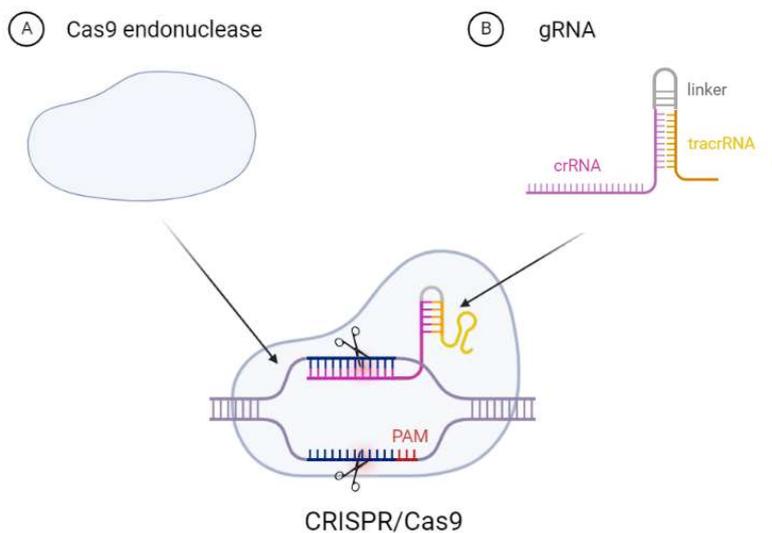
In both mouse and human, *Gata2/GATA2* and *Gata3/GATA3* code for TFs that are important TE markers regulating the first lineage segregation<sup>67</sup>. It has been reported that *GATA2* and *GATA3* operate downstream of the Hippo pathway during first lineage segregation in the mouse embryo<sup>19</sup>. However, it has not been proven that the Hippo pathway is conserved across mouse and human species<sup>67</sup>. Additionally, it remains unsure if functions of *GATA2* and *GATA3* are conserved across mouse and human species<sup>67</sup>. Home *et al.* showed that knockdown/knockout of *GATA3* in mouse embryos partially impaired morula to blastocyst transformation<sup>68,69</sup>. Moreover, another study of Home *et al.* in 2017 revealed that *Gata2* function is not necessary for blastocyst formation, as *Gata2* conditional KO mouse embryos could reach the blastocyst stage<sup>69</sup>. Therefore, it is expected that *Gata3* *-/-* embryos will mostly adopt the morula phenotype, whereas it is foreseen that *Gata2* *-/-* embryos will still be able to reach the blastocyst stage. Besides, functional redundancy has been demonstrated

in mice between GATA2 and GATA3, as GATA2 expression can restore a GATA3 KO mouse embryo's phenotype<sup>68</sup>. This was supported by the study of Home *et al.*, as *Gata2/3* double KO (DKO) in mouse embryos resulted in dysfunctional blastocyst formation<sup>69</sup>. Furthermore, in mouse, both GATA2 and GATA3 induce expression of TE specific genes such as *Prl3d1* and *Prl2c2* *in vivo* and *in vitro*<sup>70</sup>. In addition, GATA2 and GATA3 are expressed in TGCs in the mouse embryo and carry out functions in the mouse placenta<sup>70</sup>. Finally, Krendl *et al.* revealed that both GATA2/3 and TFAP2A/C are involved in the differentiation of human ESCs (hESCs), towards TE stem cells, by repression of pluripotency and the induction of expression of placenta associated genes<sup>71</sup>. These are the first indications of the potential correlation between GATA2/3 and TFAP2A/C in the formation, specification and maintenance of the TE in human embryos<sup>71</sup>.

## 1.6 CRISPR/Cas9

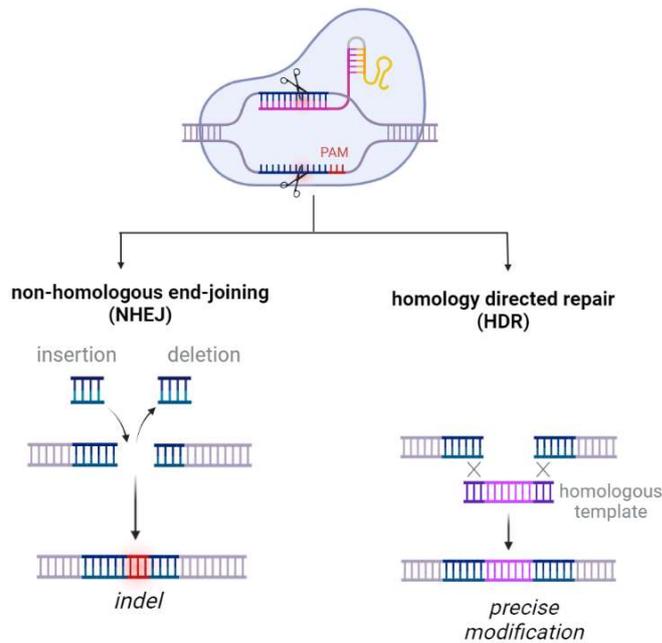
The recent advancements in clustered regularly interspaced short palindromic repeats (CRISPR)/ CRISPR-associated protein 9 (Cas9) (CRISPR/Cas9) gene editing has allowed researchers to investigate the functions of numerous genes which are important during preimplantation development. Therefore, we will make use of this gene editing system to elucidate the function of our abovementioned genes of interest.

CRISPR/Cas9 is a gene-editing technology derived from a natural defense mechanism of bacteria against invading viruses, which makes use of a Cas9 protein, a DNA endonuclease capable of inducing double stranded breaks (DSBs) in target DNA<sup>72</sup>. Additionally, a guide RNA (gRNA), consisting of the crRNA and tracrRNA duplex, will ensure guidance of the Cas9 towards the target DNA, through complementary binding<sup>72</sup>. Within the gRNA, the crRNA will bind to the DNA target site, which will be 18-20 base pairs (bp) in length, whereas the tracrRNA acts as a docking site for the Cas9 endonuclease<sup>73</sup>. Via this mechanism, the crRNA is able to guide the Cas9 endonuclease towards the target site, while the tracrRNA component acts as a connecting mechanism between these two subparts<sup>74</sup>. Furthermore, the target site in the DNA must be situated next to a protospacer adjacent motif (PAM), which is present at the 3' end of the DNA strand opposing the DNA strand binding the gRNA<sup>74</sup>. This is necessary as the Cas9 endonuclease must recognize the PAM sequence before a double-stranded break (DSB) can be introduced<sup>74</sup>. The bacterium species from which the Cas9 protein originates determines the PAM sequence<sup>72,74</sup>. The most well-known Cas9 molecule (which will also be used for the research of this master's thesis) originates from *Streptococcus Pyogenes*, relying on a 5'-NGG-3' (N being either A, T, C or G) PAM sequence<sup>74</sup>. The Cas9 endonuclease will cause a double-stranded break (DSB) on the target site, 3 bp upstream of the PAM sequence, following the binding of the crRNA to the target site on the DNA<sup>73</sup>. The structure of CRISPR/Cas9 is displayed in **Figure 7**.



**Figure 7: CRISPR/Cas9 structure.** The Cas9 endonuclease (A) is coupled to a crRNA-tracrRNA duplex (gRNA) (B) that will guide this endonuclease towards the target region (blue). A protospacer adjacent motif (PAM) motif (red) is required next to the target site. Double stranded breaks (DSB) will be caused 3 bp upstream of the PAM sequence.

The DSB generated by the CRISPR/Cas9 gene-editing system can be repaired by cell's endogenous repair mechanisms <sup>72</sup>. In mammals, two repair pathways are most prevalent in the repair of DSBs, them being, non-homologous end-joining (NHEJ) and homology-directed repair (HDR) <sup>75</sup>. The mechanism of NHEJ and HDR are displayed in **Figure 8**. NHEJ is an error-prone repair pathway which can result in small insertions or deletions (indels) at the cut site <sup>76</sup>. This way, out-of-frame or frameshift mutations, meaning they can cause a shift in the open reading frame (ORF), can lead to the formation of a premature stop codon and transcription of a mutant mRNA <sup>77</sup>. Subsequently, the process of nonsense-mediated decay (NMD) will elicit the degradation of the mutant mRNA <sup>77</sup>. In such manner, a knock-out (KO) of the target gene can be generated by creating loss-of-function mutations, allowing us to study the function of the target gene <sup>77</sup>. Also, a DSB can be repaired by HDR <sup>76</sup>. The ends flanking the DSB are rejoined by HDR in the presence of a donor DNA template homologous to the target site <sup>76</sup>. The latter repair pathway allows precise gene corrections or gene knock-ins (KI) <sup>76</sup>. Most repair pathways are active in the majority of cells, yet for a variety of reasons, most mammalian cells prefer to use NHEJ to repair a DSB <sup>75,78</sup>. While HDR is active only during the S/G2 phase of the cell cycle, NHEJ is active throughout the entire cell cycle, in addition to the faster process of NHEJ <sup>75,78</sup>. As a result, even if a DNA template homologous to the target site is present, NHEJ may be the preferred pathway to repair the DSB <sup>75</sup>. To overcome this, the HDR pathway can be favored by inhibiting NHEJ (with inhibitors such as Scr7 or Nu7026) or enhancing HDR activity (with molecules such as L755507 and Brefeldin A) <sup>75,79</sup>. Besides repair by NHEJ and HDR, the DSB can also be repaired by microhomology-mediated end-joining (MMEJ) <sup>76</sup>. If the DSB is repaired by MMEJ, identical micro homologous sequences (>2 bp) that surround the DSB are annealed, followed by the repair of the DSB <sup>76</sup>. The micro homologous sequences that border the cleavage site determine how MMEJ produces deletions <sup>76</sup>. MMEJ makes use of recurrent short sequence patterns surrounding the location of the DSB which allows for more precise repair outcome predictions as this process will result in repeatable mutations <sup>80</sup>. DSB repair via MMEJ can also lead to frameshift mutations, resulting in the formation of a premature stop codon <sup>76</sup>. Thus, repair of the DSB by NHEJ or MMEJ is desired for our applications.



**Figure 8: Modes of DSB repair by CRISPR/Cas9.** NHEJ is an error prone pathway for repairing DSBs, which can lead to indels at the on-target region. Based on a DNA template, which contains regions similar to the regions flanking the DSB, HDR rejoins the ends flanking the DSB.

### 1.6.1 Limitations

CRISPR/Cas9 has proven to be a very effective tool for studying gene function in developmental studies. Yet, there are multiple limitations to the use of CRISPR/Cas9 gene editing in the germline that need to be overcome.

First, in addition to the target site, CRISPR/Cas9 can also bind to homologous sequences in the genome, referred to as off-target sites, and usually contain one or more bp mismatches<sup>81</sup>. In comparison to on-target areas, off-target regions have a propensity for more pronounced binding and cleavage when there are fewer mismatches<sup>81</sup>. Evidently, for our applications, off-target effects are not feasible as they could cause the displayed phenotype rather than an on-target frameshift mutation. This way, no solid conclusions can be drawn in terms of gene function. Therefore, the occurrence of off-target mutations needs to be reduced as much as possible. A lot of *in silico* prediction tools (such as Benchling) have been created to design a crRNA that targets a specific region in a gene, while minimizing the chance of off-target effects<sup>81</sup>. Additionally, the CRISPR/Cas9 delivery method can also affect the prevalence of off-target effects<sup>82</sup>. By making use of plasmid transfection-mediated CRISPR/Cas9 delivery, subsequent high enzyme concentrations will boost aspecific binding and cleavage at off-target sites<sup>82</sup>. Thus, the use of purified recombinant Cas9 protein in ribonucleoprotein (RNP) complexes may increase specificity by decreasing enzymatic exposure time and off-target effects<sup>82</sup>. Furthermore, the use of the CRISPR/Cas9 RNP complex increases the chance of inducing KO before the first cell division occurs in the zygote, as transcription from plasmids is not necessary prior to the editing<sup>81,82</sup>.

Moreover, in this context, the occurrence of genetic mosaicism during preimplantation development is a hurdle that needs to be overcome<sup>83</sup>. Genetic mosaicism is the phenomenon where certain cells of the embryo are edited, while other ones remain unedited within the same embryo, and is caused when CRISPR/Cas9 mediated gene editing precedes the first cell division<sup>83</sup>. The emergence of mosaicism, which may impede phenotypic analysis and interpretation of developmental behavior of the embryos, is one of the main challenges regarding CRISPR/Cas9 germline genome editing<sup>48</sup>. It is expected that mosaicism will be reduced when the CRISPR/Cas9 complex will be delivered as early as possible before the first cell division<sup>83</sup>.

Additionally, the CRISPR/Cas9 gene editing system can induce large chromosomal deletions or complex genomic rearrangements such as inversions, large insertions, translocations and (partial) chromosome loss <sup>74,84-86</sup>. Likewise, it can potentially occur that the other chromosome serves as a repair template for the on-target site containing the DSB, which is called inter-homolog homologous repair (IH-HR) with crossover at the on-target site <sup>74,84-86</sup>. These phenomena can lead to so-called loss of heterozygosity (LOH), where one allele is lost, leading to the loss of genetic contribution of one parent in diploid cells <sup>74,87</sup>. It must be noted that large chromosomal deletions may go unnoticed when looking into their potential occurrence, as genotyping the genomic region targeted by the CRISPR/Cas9 system implies the PCR amplification of a small region (400-600 bp) surrounding the on-target site <sup>87</sup>. The annealing site of one of both primers will be lost if deletions caused by CRISPR/Cas9 are larger than these fragments in either direction <sup>87</sup>. As a result, only one allele is amplified, giving the impression that editing was unsuccessful or that only one homozygous event took place at the on-target site <sup>84-86,87</sup>. A possible technique to check whether large chromosomal deletions occurred as a result of CRISPR/Cas9 genome editing, is long-range PCR (capable of amplifying regions of up to 30 kilo bases (kb)), which can be used for genotyping a larger region surrounding the on-target locus <sup>88</sup>. However, long-range PCR shows some limitations, such as the time needed to perform PCR reactions, and the low fidelity of the DNA polymerase, which restricts its use in this context <sup>88</sup>. Additionally, shallow whole genome sequencing (WGS) can provide insight in potential chromosomal aberrations caused by CRISPR/Cas9 activity <sup>89</sup>.

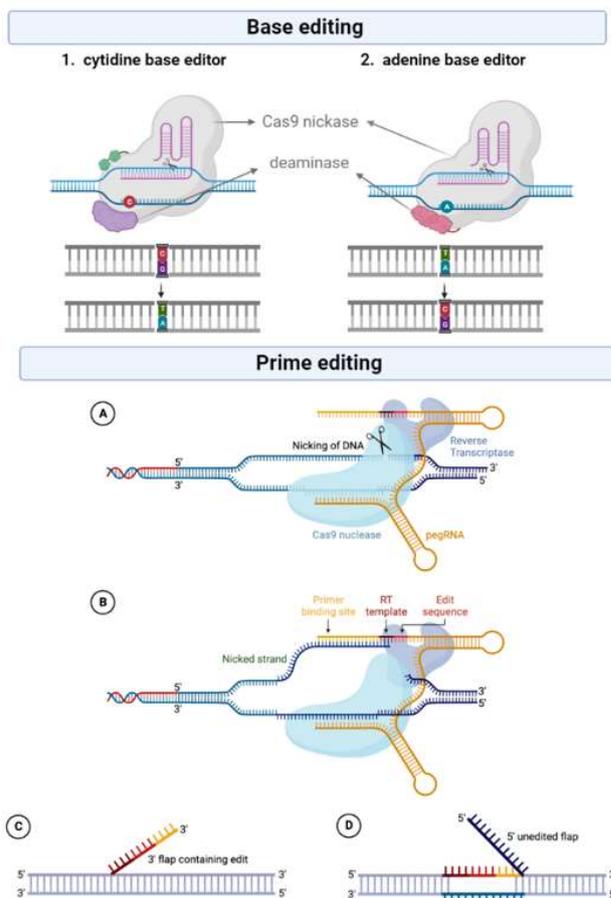
### 1.6.2 *Alternative approaches*

Even though CRISPR/Cas9 has proven to be a useful tool for germline genome editing, the aforementioned phenomena highlight that the genome editing system still has a number of drawbacks. Therefore, more advanced CRISPR/Cas9 variants that do not rely on DSB formation, such as base editing and prime editing, can be researched in the context of germline genome editing <sup>74</sup>. The use of base editing and prime editing variants result in more precise on-target editing and reduce off-target effects <sup>74</sup>. The CRISPR/Cas9 prime editor (**Figure 9B**) consists of a Cas9 nickase, coupled with a reversed transcriptase and a prime editing gRNA (pegRNA), capable of identifying the target sequence and of encoding new genetic information which replaces the target sequence <sup>90</sup>. By making use of a Cas9 nickase, the generation of a DSB is avoided <sup>90</sup>. Instead, the complementary DNA strand is nicked three bases upstream of the PAM site <sup>90</sup>. As a consequence, a DNA flap with a 3' OH group is exposed, which binds to the primer binding site (PBS) of the RNA template, which will act as a primer for the reversed transcriptase <sup>90</sup>. Subsequently, by copying the edit sequence of the pegRNA, the 3' flap will be extended <sup>90</sup>. The prime editing system is capable of making any base pair change, as well as small indels, and is hence suited for both gene correction and the generation of KOs <sup>74,91</sup>. Although recent experiments with prime editing seemed promising, optimization of this gene editing system remains necessary to further investigate its function in germline genome editing <sup>74</sup>. Furthermore, base editors (**Figure 9A**) also contain a Cas9 nickase, a gRNA and a deaminase, and are capable of chemically modifying the DNA base pairs, resulting in either a C:G to T:A (cytidine base editor, **Figure 9A.1**) or a A:T to G:C (adenine base editor, **Figure 9A.2**) base pair transition, thus able to generate gene KOs and to correct a certain bp mutation <sup>74,92</sup>. Base editors are capable of mediating several transition mutations without generating DSBs, using homologous repair templates or depending on the HDR pathway <sup>93</sup>. Therefore, in theory, the base editors have an advantage over the current CRISPR/Cas9 gene editing system since also no DSBs are generated, yet

the influences on the chromosomal constitution of the germline genome still need to be researched extensively <sup>74</sup>.

Prime editing has already been successfully performed in human embryonic stem cells (hESCs), as the prime editor was able to introduce all types of nucleotide substitutions, and small insertions and deletions <sup>94</sup>. Interestingly, it was also reported that a prime editor's reverse transcriptase domain does not result in off-target mutations in the genome in hESCs <sup>94</sup>. Moreover, base editors have been applied successfully in both mouse and human embryos and were proven effective <sup>95,96</sup>. Therefore, both base and prime editing show potential for clinical applications rather than the original CRISPR/Cas9 gene editing system, as they do not rely on DSB generation <sup>74</sup>. Additionally, both mechanisms display a reduced chance for off-target effects, and therefore show improved safety for clinical applications <sup>74,90,97</sup>.

In summary, both base and prime editing seem promising for editing the germline genome in the future, yet extensive research must be done to look into the safety of use of these gene editing tools. Hence, we will remain making use of the CRISPR/Cas9 gene editing system to perform experiments of germline genome editing.



**Figure 9: Base and prime editing mechanisms. (A) Base editing.** The Cas9 nickase is coupled to a guide RNA (gRNA), which will guide the nickase towards the target site. At the target site, the deaminase will ensure the base pair conversion (either C:G → T:A or T:A → C:G). **(B) Prime editing.** The Cas9 nickase nicks the target DNA. Next, reverse transcription follows the hybridization of the primer binding site (PBS) to the PAM strand. Finally, ligation and mismatch repair take place after the hybridization of the DNA strands and the flap cleavage.

## 1.7 Scientific research objectives

By creating KO mouse embryos of our genes of interest (*Gata3*, *Gata2*, *Tfap2c* and *Tfap2a*) we can gain insight into the molecular mechanisms involved in formation, specification and maintenance of the TE. GATA3 and TFAP2C (and their isoforms GATA2 and TFAP2A) are important TE markers, yet their function in the early mouse embryo remains unelucidated. The results of this research could provide more insight into certain causes of implantation failure after ART, potentially leading to improved media culture conditions of embryos created via ICSI or IVF, preceding embryo transfer. Therefore, in this master's dissertation we will specifically research following objectives:

- Create 100% *Gata3* *-/-*, *Gata2* *-/-*, *Tfap2c* *-/-* or *Tfap2a* *-/-* mouse embryos and assess the effect of the gene KO on blastocyst formation.
- Investigate the function of early embryonic genes such as *Gata3*, *Gata2*, *Tfap2c* and *Tfap2a* in specification, formation and maintenance of the TE in mouse embryos.

In the future, the consequences of combinatorial loss of *Gata2/3* and *Tfap2a/c* on mouse embryo morphology will also be assessed by creating a DKO model for both set-ups. Additionally, as important interspecies differences between mouse and human embryos have already been revealed (reviewed in <sup>4</sup>), the function of these genes will also be researched directly in human embryos on the long term.

## 2. Materials and methods

### 2.1 crRNA and primer design

First, the genomic target region in the genes of interest (GOI) was selected. As a general rule, the first exon could not be targeted, as there is a risk for so-called exon skipping, which is the phenomenon where an alternative first exon is used, as a result of the formation of an early stop codon in the first exon <sup>98</sup>. The designed crRNA could also not target the last exon or the last 55 bp of the second-last exon, as it is possible that the mRNA product of the KO gene would not be degraded by the nonsense-mediated pathway (NMD) anymore, resulting in a potential functioning protein product <sup>77</sup>. Considering the abovementioned criteria, two possible crRNAs were designed for each GOI, by making use of the online *in silico* tool Benchling. Benchling contains built-in algorithms to assess the maximal on-target and off-target potential of the designed crRNAs. Several crRNA designs with a high on- and off-target score were selected for further analysis. To be clear, crRNA designs with high off-target scores have a lower chance of causing off-target effects and were therefore desired. The online *in silico* tool InDelphi was used to analyze a subset of pre-designed crRNAs (with high on-target and off-target scores) to determine whether the designed crRNA is more likely to cause in-frame or out-of-frame mutations. In addition, most frequently predicted indels at the on-target site (accompanied by a predicted frequency), frameshift frequency, and microhomology strength were also assessed with InDelphi. Finally, two crRNAs with high on-target and low off-target potential, which would most likely generate an out-of-frame mutation, were selected for further testing. For each crRNA design, Primer3Plus was used to design two pairs of primers for each respective target site, to make amplification of this target region possible.

## 2.2 Mouse zygote collection

Female B6D2 mice (between 6 and 12 weeks of age) underwent ovarian stimulation by intraperitoneal (IP) injection of 5 IU pregnant mare serum gonadotropin (PMSG, MSD animal health) followed by the injection of 5 IU ml human chorion gonadotropin (hCG, MSD animal health) 48 hours later. Female mice were mated overnight immediately following hCG injection, and 20 hours later, the mice were sacrificed by cervical dislocation. Ovaria and oviducts were dissected and placed in 2 mL KSOM-HEPES medium (in-house made) enriched with 0.4% bovine serum albumin (HEPES BSA). For all following steps, used culture media were covered with OVOIL™ heavy embryo culture oil (#10174, Vitrolife) to counteract evaporation of the culture medium. First, cumulus oocytes complexes were harvested from the ampullae and incubated in drops of hyaluronidase (200 IU/mL, #H3506, Sigma-Aldrich) for 3-5 minutes, to remove the cumulus cells from the zygotes. Additionally, mechanical dissociation was performed by pipetting up and down with a sterile Pasteur Pipette. Zygotes were washed in 25 µL drops of HEPES BSA medium and KSOM low glucose medium (in-house made) enriched with 0.4% BSA (KSOM-BSA). Finally, the zygotes were transferred to 25 µL drops of KSOM-BSA and incubated at 37°C (5% O<sub>2</sub> and 6% CO<sub>2</sub>). Mouse embryos were cultured in KSOM-BSA up until the 8-cell stage and were transferred to CookBlasto® blastocyst medium (#K-SIBM-50, Cook Medical) 68-72 hours after the hCG administration and incubated at 37°C (5% O<sub>2</sub> and 6% CO<sub>2</sub>).

## 2.3 RNP preparation for electroporation/nucleofection

Upon arrival, crRNA/tracrRNA oligos were resuspended in nuclease free water to a final concentration of 100 µM. To prepare the ribonucleoprotein (RNP) complex for electroporation purposes, the designed crRNA (#1072532, IDT) and tracrRNA (#1072532, IDT) were mixed in equimolar concentrations in a sterile microcentrifuge tube to achieve a 50 µM concentration of the crRNA-tracrRNA duplex. This was incubated at 95°C for 5 minutes to allow crRNA-tracrRNA duplex formation. Next, Cas9 protein (#1081058, IDT) was added to the crRNA-tracrRNA duplex (**Table 1**) and was incubated at RT for 15 minutes to allow RNP complexation. Finally, this complex was diluted in opti-MEM buffer until a final volume of 50 µL (#31985-062, Thermo Fisher Scientific) and stored at -80°C until use.

Preceding nucleofection, RNP complexes were prepared and for each nucleofection reaction, the crRNA-tracrRNA duplex and Cas9 (#1081058, IDT) were diluted in PBS 1X (#14190-094, Thermo Fisher Scientific) until a final volume of 4.72 µL was reached (**Table 2**). Next, the RNP complexes were incubated at RT for 20 minutes and subsequently stored at 4°C before use.

**Table 1:** Composition of CRISPR/Cas9 complex for electroporation.

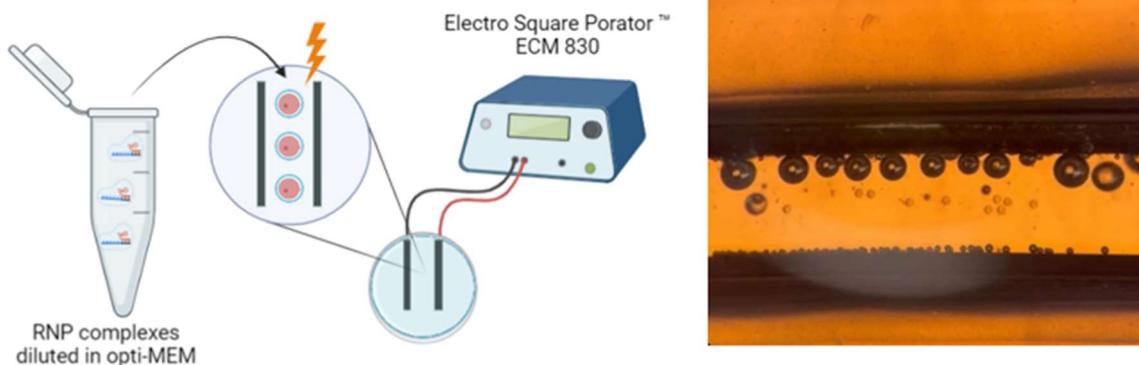
Component	Concentration
crRNA-tracrRNA duplex (50 µM)	3 µM
Alt-R Cas9 enzyme (#1081058, IDT) (61 µM)	1.2 µM
Opti-MEM (#31985-062, Thermo Fisher Scientific)	/

**Table 2:** Composition of CRISPR/Cas9 complex for nucleofection.

Component	Concentration
crRNA-tracrRNA duplex (50 $\mu$ M)	21 $\mu$ M
Alt-R Cas9 enzyme (61 $\mu$ m stock)	1.8 $\mu$ M
PBS 1X (#14190-094, Thermo Fisher Scientific)	/

#### 2.4 Electroporation: CRISPR/Cas9 delivery in mouse zygotes

CRISPR/Cas9 RNP complexes were delivered to mouse zygotes via electroporation (**Figure 10**) using the Electro Square Porator™ ECM 830. The RNP complexes were delivered at the zygote stage, as this reduces the possibility of genetic mosaicism<sup>83</sup>. Up until the washing step of the embryos in HEPES BSA (in-house made), routine zygote collection was carried out. Following, zygotes were washed 5 times in opti-MEM medium (#31985-062, Thermo Fisher Scientific) and aligned between the two electrodes of a BTX Microslide™ (#450104, BTX®) electroporation dish in 5  $\mu$ L RNP solution (prepared as mentioned in section 2.3). Electroporation was performed with the parameters presented in **Table 3**. Next, zygotes were retrieved and washed in HEPES-BSA before they were washed and cultured in KSOM-BSA (in-house made) at 37°C (5% O<sub>2</sub> and 6% CO<sub>2</sub>).



**Figure 10: Electroporation of mouse zygotes for delivery of CRISPR/Cas9 components. (A)** Electroporation set-up for mouse zygotes. **(B)** Microscopic image taken of electroporation reaction.

**Table 3:** Electroporation parameters used for CRISPR/Cas9 delivery in mouse oocytes.

Parameter	Value
Voltage	30 V
Number of pulses	2
Pulse length	2 ms
Pulse interval	100 ms

## 2.5 mESC culture prior to nucleofection

Stable, naïve mESCs (R1WT) were thawed minimally one week before nucleofection. Mouse ESCs were cultured in serum-leukemia inhibitory factor (LIF) conditions (**Table 4**) on a 0.1% gelatin (in-house made) coated 100 mm x 20 mm Style dish for Cell Culture (#430167, Corning®) and incubated at 37°C (5% O<sub>2</sub> and 6% CO<sub>2</sub>).

**Table 4:** Serum-leukemia inhibitory factor (LIF) medium composition.

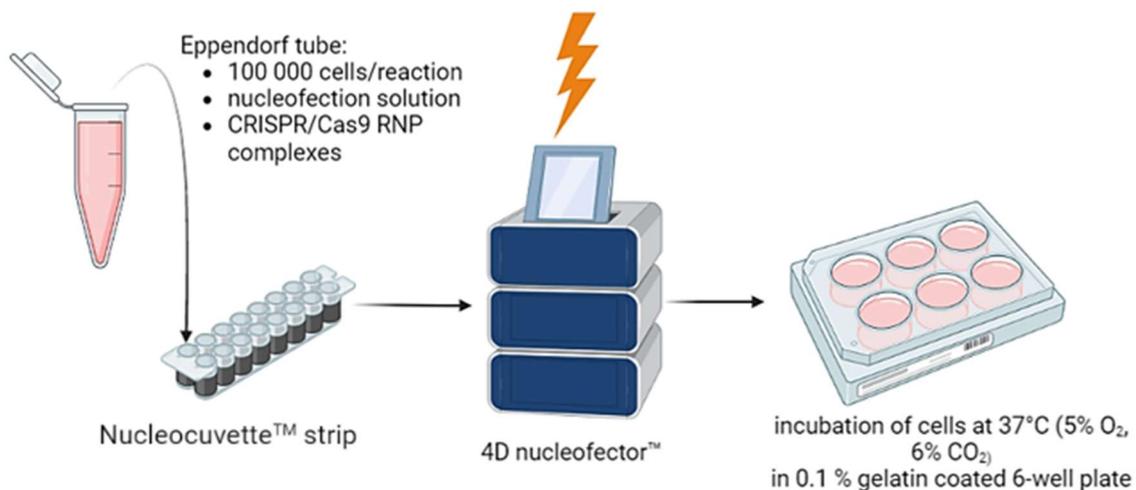
Compound	Cat. No.	Company	Concentration
Fetal bovine serum (FBS) (15%)	#10270106	Thermo Fisher Scientific	1/8 (vol/vol)
Penicilin/streptomycin (pen strep)	#15140122	Thermo Fisher Scientific	1/100 (vol/vol)
L-glutamine	#25030024	Thermo Fisher Scientific	1/100 (vol/vol)
Sodium pyruvate	#11530396	Thermo Fisher Scientific	1/100 (vol/vol)
Minimum essential medium non-essential amino acids (MEM NEAA)	#11140050	Thermo Fisher Scientific	1/100 (vol/vol)
β-mercaptoethanol	#31350010	Thermo Fisher Scientific	1/600 (vol/vol)
Mouse leukemia inhibitory factor (mLIF)	#78056	Stem Cell Technologies	1/10344.8 (vol/vol)
Dulbecco's modified Eagle medium (DMEM)	#62248	Thermo Fisher Scientific	1/1.2 (vol/vol)

Next, mESCs were splitted 1:15 when 70% confluency was reached. First, a 100 mm x 20 mm Style dish for Cell Culture (#430167, Corning®) was covered with 2 mL 0.1% gelatin (in-house made) and incubated for 20-30 minutes. Next, mESCs were washed 3 times with PBS 1X (#14190-094, Thermo Fisher Scientific) and incubated with 2 mL 0.05% trypsin (#25300, Thermo Fisher Scientific) for 2-3 minutes. Subsequently, 0.05% trypsin (#25300, Thermo Fisher Scientific) was inactivated by adding 2 mL serum-LIF medium. The cell suspension was centrifuged for 3-4 minutes at 300 relative centrifugal force (rcf) and the supernatant was discarded from the pellet containing the single mESC cells. The pellet was resuspended in 3 mL serum-LIF medium, and 0.2 mL of this suspension was added to 8 mL of fresh serum-LIF medium and plated on the 0.1% gelatin coated 100 mm x 20 mm Style dish for Cell Culture (#430167, Corning®). The R1WT naïve mESCs were intended to be split minimally once prior to nucleofection to ensure that cells were stable enough to go through the nucleofection process.

## 2.6 Nucleofection procedure

Nucleofection was carried out using the 4D nucleofector™ (Lonza) in combination with the Amaxa™ P3 primary Cell 4D nucleofector X kit. First, the Alt-R Cas9 Electroporation Enhancer (#1075916, IDT) was resuspended in 100 µM Tris-EDTA (TE) buffer (# 93283-100ML, Sigma Aldrich). Next, regular splitting was performed (as described in section 2.5) and cells were washed in PBS 1X (#14190-094, Thermo Fisher Scientific) prior to counting the mESCs using a Neubauer counting chamber. A total amount of 100 000 cells was

included per reaction. Nucleofection solution was prepared by adding 13.94  $\mu\text{L}$  P3 primary cell solution to 3.06  $\mu\text{L}$  of supplement 1 for 1 reaction. Subsequently, 4,20  $\mu\text{L}$  of the crRNA-tracrRNA-Cas9 RNP complex was supplemented with 0,80  $\mu\text{L}$  of 100  $\mu\text{M}$  Alt-R Cas9 electroporation enhancer (#1075916, IDT). Next, Eppendorf tubes containing 100 000 cells were centrifuged at 300 rcf for 4 minutes, supernatant was discarded and 22  $\mu\text{L}$  of the nucleofection solution was added to the cells. Next, 20  $\mu\text{L}$  of the RNP-cell suspension was added to a well of the 16-well Nucleocuvette™ strip (#PDH-2104, Lonza) and cells were nucleofected with nucleofection program DN100 on the 4D nucleofector™ (Lonza). After nucleofection, 80  $\mu\text{L}$  of prewarmed serum-LIF medium (see section 2.5) was added to each well of the Nucleocuvette™ strip and the complete content of each well of the 16-well Nucleocuvette™ strip (#PDH-2104, Lonza) was added to one well of a 6-well plate for cell culture. Lastly, cells were incubated for 48 hours at 37 °C (5% O<sub>2</sub>, 6% CO<sub>2</sub>) on previously incubated gelatin-coated plates. Two days after nucleofection, DNA was extracted from all the cells located in one well (see section 2.8.2). **Figure 11** displays the used set-up for nucleofection of the naïve mESCs with CRISPR/Cas9 RNP complexes.



**Figure 11:** Set-up for nucleofection of naïve mouse embryonic stem cells (mESCs) with CRISPR/Cas9 ribonucleoprotein (RNP) complexes.

## 2.7 Immunohistochemistry and imaging

Embryos were fixed by incubating them in 4% paraformaldehyde (PFA, in-house made) for 30 minutes at RT on an orbital shaker platform, followed by several washing steps in PBS 1X buffer (**Table 5**). Upon permeabilization in PBT 0.5% (**Table 5**) for 20 minutes at RT, blocking was performed by immersing the embryos in blocking solution (**Table 5**) (1h, RT). Overnight, embryos were incubated with the primary antibodies (**Table 6**) diluted in blocking solution (1/500 vol/vol). The following day, embryos were washed three times in PBT 0.1% (**Table 5**) before incubation with the secondary antibodies (**Table 7**) and DAPI (#62248, Thermo Fisher Scientific) diluted in blocking solution for one hour at RT. DAPI (#62248, Thermo Fisher Scientific) was used to visualize the chromosomes. The secondary antibodies were used in a dilution of 1/500 (vol/vol) whereas the DAPI was used in a dilution of 1/50

(vol/vol) in the blocking solution. Next, the embryos were washed in blocking solution and the Willco-dish® glass bottom imaging chamber (#GWST-5040, WillCo Wells BV) was prepared by placing drops of PBS 1X on the imaging chamber covered with OVOIL™ heavy embryo culture oil (#10174, Vitrolife). Finally, single embryos were placed in a drop of PBS 1X and visualized with the Zeiss LSM9000 confocal microscope.

**Table 5:** Composition of media prepared for immunofluorescent staining.

Medium	Composition
PBS 1X	Embryo transfer water (ETW, #W1503-500ML, Sigma Aldrich) containing PBS 10X (#AM9624, Thermo Fisher Scientific)
PBT 0.1%	PBS 1X supplemented with 0.1% Triton X-100 (#T8787-50ML, Sigma Aldrich)
PBT 0.5%	PBS 1X supplemented with 0.5% Triton (#T8787-50ML, Sigma Aldrich)
Blocking solution	PBT 0.1% supplemented with fetal bovine serum (FBS, #10270106, Thermo Fisher Scientific)

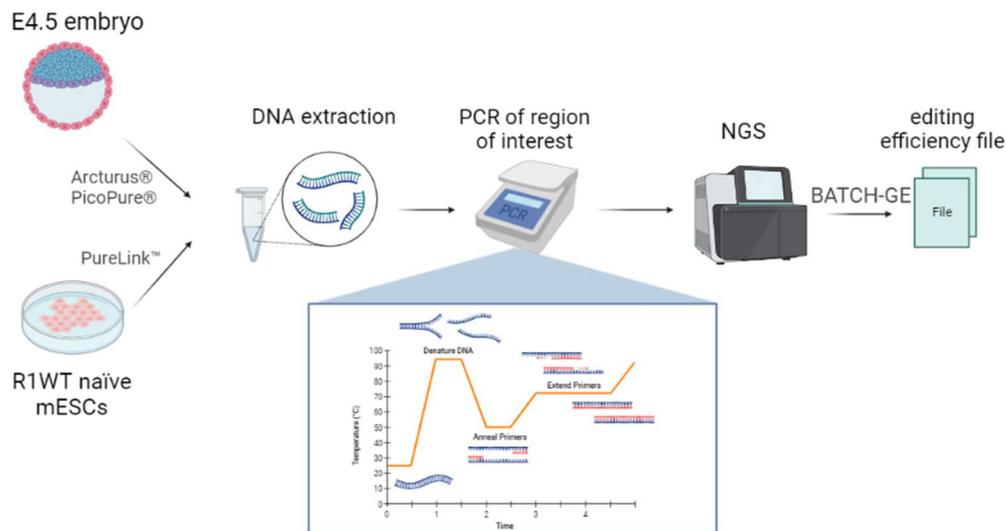
**Table 6:** Primary antibodies, alongside their origin, reference number and company, and their used concentrations.

Primary antibody	Origin	Cat. No.	Company
GATA3	Goat	#AF2605	R&D systems
GATA2	Mouse	#WH0002624M1	Sigma Aldrich
TFAP2A	Rabbit	#AB108311	Abcam
TFAP2C	Mouse	#SC-12726	Santa Cruz
CDX2	Rabbit	#AB-76541	Abcam

**Table 7:** Secondary antibodies, alongside their origin, reference number and company, and wavelength at which they are excited.

Secondary antibody	Origin + target	Cat. No.	Company	Wavelength
Alexa fluor 647	Donkey anti-goat	#A21447	Thermo Fisher Scientific	647 nm
Alexa fluor 488	Donkey anti rabbit	#A21206	Thermo Fisher Scientific	488 nm
Alexa fluor 594	Donkey anti-mouse	#AB150108	Abcam	594 nm

## 2.8 DNA extraction and genotyping



**Figure 12:** Workflow of DNA extraction and genotyping of single embryos and naïve mESCs.

### 2.8.1 DNA extraction of embryos

The Arcturus® PicoPure® DNA Extraction Kit (#KIT0103, Thermo Fisher Scientific) was used to extract the DNA from single CRISPR/Cas9 targeted, scrambled, and control embryos (morula or blastocyst phenotype). The DNA extraction solution was first prepared by dissolving the contents of one proteinase K vial in 155  $\mu$ L reconstitution buffer. Single embryos were placed in 10  $\mu$ L of DNA extraction solution and incubated at 65°C for 3 hours, followed by incubation of 10 minutes at 95°C to inactivate the proteinase K.

### 2.8.2 DNA extraction of naïve mESCs

The PureLink™ Genomic DNA Mini Kit (#2463685, Thermo Fisher Scientific) was used to perform whole well DNA extraction of R1 naïve mESCs after nucleofection. For each well, cells were splitted as described above (see section 2.5). To each Eppendorf tube containing the cells from one reaction, 20  $\mu$ L proteinase K was added, followed by 20  $\mu$ L RNase A. Next, 200  $\mu$ L PureLink Genomic Lysis/Binding buffer was added and the Eppendorf tube and was vortexed to obtain a homogenous solution. The sample was incubated at 55°C for 10 minutes to promote protein digestion. Following, 200  $\mu$ L of 100% ethanol was mixed with the lysate and the mixture was transferred to a PureLink Spin Column enclosed by a collection tube. The column was centrifuged at 10 000 rcf for 1 minute at RT and the content of the collection tube was discarded. The same process was repeated with 500  $\mu$ L Wash Buffer 1 (prepared with ethanol). Next, 500  $\mu$ L Wash Buffer 2 was added to the column and the column was centrifuged for 3 minutes at maximum speed (12 000 rcf) at RT. Finally, 50  $\mu$ L of PureLink Genomic Elution Buffer was applied to the column, and the column was incubated for 1 minute at RT and centrifuged for 1 minute at maximum speed at RT. The purified DNA extract was collected in the collection tube after this step.

### 2.8.3 PCR reactions following DNA extraction followed by NGS.

PCR reactions were done on the DNA extracts of the single embryos or the naïve mESCs to allow amplification of the region surrounding the on-target site. Five µL of KAPA2G Robust Hot start Ready Mix (#KK5701, Sigma-Aldrich) (premade mixture containing DNA polymerase, DNTPs, stabilizers and MgCl) was added to a DNA low binding tube for PCR, followed by 2.5 µL of working solution of primers (2 µM) (see **Supplementary table 9**) and 2.5 µL of DNA extract. Continuously, PCR reactions were started by either following the PCR program displayed in **Table 8** (Touchdown or traditional PCR) or **Table 9**. After PCR reactions, 5 µL nuclease free water (NFW) was added to the samples, and samples were loaded on a fragment analyzer (FA). FA is a technique allowing for the separation of DNA fragments by capillary electrophoresis, followed by sizing of the fragments by comparison to an internal standard <sup>117</sup>. This was performed as a quality control step for subsequent NGS analysis. If the samples contained a minimum concentration of 1 ng/mL of the amplicon corresponding with our region of interest, they could be incorporated for next generation sequencing (NGS, Illumina™ Miseq). Data retrieved from NGS of the CRISPR/Cas9 targeted embryos was analyzed with BATCH-GE, an in-house developed software tool which detects and report indel mutations, alongside other precise genome editing events, and calculates the corresponding mutagenesis efficiencies <sup>99</sup>. This techniques provides insight into the mutations caused by CRISPR/Cas9 gene editing. However, a 5% error margin is taken into account, and therefore, mutations that occur at a relative frequency lower than 5% were not considered. NGS data was also analyzed with the Integrative Genomics Viewer (IGV), which is a tool for visualizing NGS data sets by mapping the reads obtained from bridge amplification in NGS against the reference genome <sup>100,118</sup>.

**Table 8:** Parameters of traditional PCR program, duration = hh:mm:ss

Temperature	Duration	Number of cycli
95°C	0:04:00	1
95°C	0:00:20	12
62°C	0:00:15	
72°C	0:01:00	
94°C	0:00:40	
50°C	0:00:40	24

**Table 9:** Parameters of PCR program 2, duration = hh:mm:ss

Temperature	Duration	Number of cycli
95°C	00:03:00	1
95°C	0:00:15	35
60°C	0:00:10	
72°C	0:00:15	
72°C	00:01:00	1

## 2.9 Shallow whole genome sequencing

For shallow whole genome sequencing of naïve mESCs, a pellet of naïve mESCs was snap frozen in liquid nitrogen, followed by submission of this snap frozen pellet to the Centre of Medical Genetics Ghent (CMGG), to perform shallow WGS at their department using the REPLI-G™ kit. DNA from the naïve mESCs was extracted, purified, and amplified using the PicoPlex™ whole genome amplification (WGA) kit (#R30050, TaKaRa). Next, low-coverage NGS was performed using the Illumina™ NGS platform, and NGS data was analyzed to obtain the shallow WGS profiles for each sample.

## 2.10 Statistics

SPSS® Statistics software (IBM) was used for all statistical analysis, whereas all graphs were made using Excel 2016. To test the correlation between two variables, a Chi Square test was conducted. Significance level ( $\alpha$ ) was set at 0.05, meaning that if p-values below 0.05 were detected, a statistically significant result was concluded.

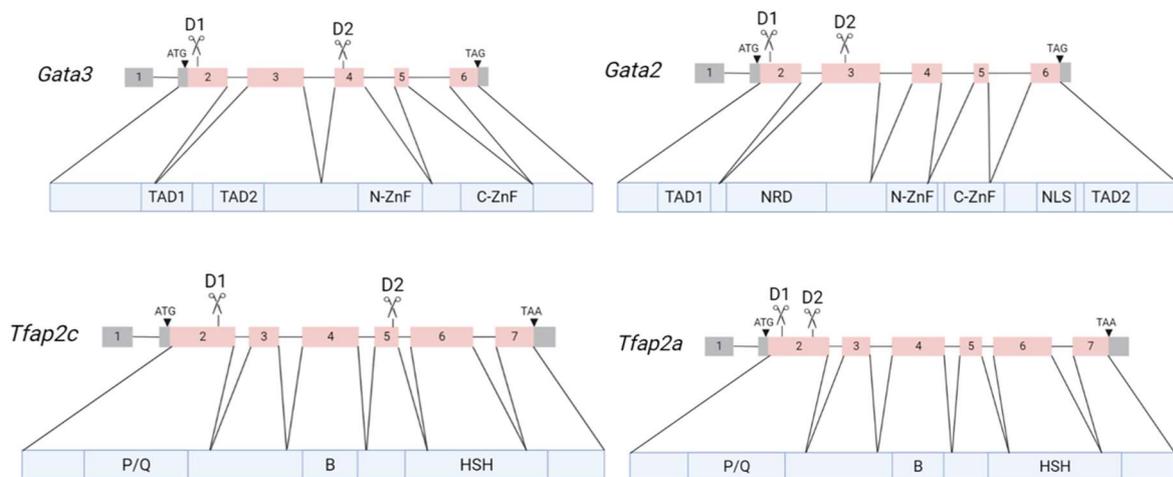
### 3. Results

#### 3.1 crRNA and primer design for each gene of interest

First, using the online *in-silico* tools Benchling and InDelphi, two crRNAs were designed for each GOI (*Gata3*, *Gata2*, *Tfap2c*, *Tfap2a*). **Table 10** displays the sequences of the two crRNA designs per gene and which exon they target, alongside their respective PAM, cleavage sites and the on-and off-target scores (predicted by Benchling). It should be mentioned that as high as possible on- and off-target scores were aimed for. **Figure 13** exhibits the gene structure of our GOIs, accompanied by the cut site of both crRNA designs per gene.

**Table 10:** crRNA design sequences per gene, their respective PAM sites and on-and off-target scores according to Benchling. ON = on-target score, OFF= off-target score

Gene	Design	Exon	Cleavage site	Sequence	PAM	ON	OFF
<b>Gata3</b>	Design 1	2	12492	GGTGAGCCACCATC ACCCCG	CGG	72.7	43.7
	Design 2	4	21170	GGGCAACCTCTACC CCACTG	TGG	76.6	41.2
<b>Gata2</b>	Design 1	2	88199562	GACACAGTAGTGGA CCATGG	AGG	75.9	63.3
	Design 2	3	88200404	GCCCCTGGGTAAAC AGACAG	AGG	67.5	50.4
<b>Tfap2c</b>	Design 1	2	172551721	GGGGTATTAGACTT GCAGAG	CGG	78.1	70.6
	Design 2	5	172554133	AGACGTGAGGAGA GTGACGT	GGG	67.7	73.6
<b>Tfap2a</b>	Design 1	2	5073	CGACGGCACCAGC AACGGGA	CGG	62.0	47.0
	Design 2	2	5391	CAGGGACTATCGG CGGCACG	AGG	62.9	62.0



**Figure 13:** Gene structure and protein coding of the murine *Gata3*, *Gata2*, *Tfap2c*, *Tfap2a* genes, accompanied by the cut sites of both crRNA designs.

Following, both crRNA designs per gene were also tested in the online *in silico* tool InDelphi, which predicts most frequently occurring insertions or deletions, and the tendency of the crRNA to generate an in-frame or out-of-frame mutation, based on the location of the cleavage site. Additionally, microhomology strength of the cleavage site can also be predicted, allowing the prediction of repair outcomes with improved accuracy if a high score was obtained. **Table 11** presents the most frequent insertion or deletion sizes at the on-target site, predicted by InDelphi, accompanied by the foreseen frameshift frequency of the crRNA designs. The most frequently occurring (*in-silico* predicted) on-target mutations for both the *Gata3* and *Tfap2a* gene targeting crRNA designs, as well as for crRNA design 1 of the *Gata2* and *Tfap2c* gene, are frameshift mutations. Besides, it was foreseen that non-frameshift mutations would result from both the *Gata2* (6 bp, 5.4%) and *Tfap2c* (3 bp, 3.3%) targeting crRNA design 2. Yet, the frequencies at which they were predicted to occur are considered rather low.

**Table 11:** Most frequent insertions or deletions for each design per gene, and their respective frameshift frequencies, predicted by InDelphi.

Gene	Design	Indel type	Predicted frequency (%)	Frameshift frequency (%)
<b><i>Gata3</i></b>	Design 1	2 bp deletion	29.9	82.4
		1 bp deletion	9.0	
		8 bp deletion	5.9	
	Design 2	1 bp deletion	12.6	82.5
		8 bp deletion	8.4	
		2 bp deletion	6.9	
<b><i>Gata2</i></b>	Design 1	7 bp deletion	39.5	89.0
		13 bp deletion	12.8	
		1 bp insertion	8.4	
	Design 2	4 bp deletion	62.4	86.9
		6 bp deletion	5.4	
<b><i>Tfap2c</i></b>	Design 1	5 bp deletion	19.1	84.7
		8 bp deletion	9.0	
		8 bp deletion	6.6	
	Design 2	5 bp deletion	38.7	83.0
		14 bp deletion	11.4	
		3 bp deletion	3.3	
<b><i>Tfap2a</i></b>	Design 1	17 bp p deletion	43.9	94.2
		5 bp deletion	17.8	
		10 bp deletion	5.1	
	Design 2	5 bp deletion	13.7	82.3
		1 bp insertion	12.2	
		7 bp deletion	8.3	

Next, using Primer3Plus, primer pairs surrounding each cleavage site were designed to enable PCR amplification and subsequent analysis of the respective amplicons. Primer pairs for each crRNA design were tested on naïve mESC DNA (WT) and analyzed with a fragment analyzer to allow for the evaluation of both their ability to amplify the region surrounding the target site (region of interest), and their specificity to this region. **Supplementary figure 1** displays the size (bp) of the amplicons corresponding with the region surrounding the target site of each crRNA per gene. **Supplementary table 9** shows the sequence of both forward and reverse primers amplifying the regions of interest.

## 3.2 Optimization of CRISPR/Cas9 components in mouse ESCs

### 3.2.1 On-target efficiencies

Before applying the CRISPR/Cas9 components directly in mouse embryos, their activity was studied in naïve mESCs, as they display translational value for editing efficiencies in mouse embryos <sup>47</sup>. Both scrambled and *Gata2*, *Gata3*, *Tfap2c* and *Tfap2a* targeting CRISPR/Cas9 RNP complexes were delivered in R1WT naïve mESCs via nucleofection. Scrambled CRISPR/Cas9 complexes contain a crRNA which has no binding site in the DNA, and therefore no editing should be detected. This group is used as an extra control group, to check for random activity of the Cas9 endonuclease and possible negative consequences of the nucleofection procedure. Continuously, DNA was extracted from the mESCs after 48-72 hours, the region surrounding the target site (on-target) was amplified using PCR and amplicons were analyzed with the Illumina™ Miseq NGS platform. Consecutively, data obtained from NGS was analyzed with BATCH-GE. **Figure 14** displays on-target editing efficiencies for the regions targeted by both crRNA designs per GOI in CRISPR/Cas9 targeted mESCs, besides editing efficiencies for the same regions in scrambled targeted mESCs (n=1). It must be noted that editing events with a relative frequency lower than 5% were omitted from the analysis, taking into account a 5% error margin for NGS. Highest frameshift efficiencies were reached with crRNA design 2 for both the *Gata3* (78%) and *Tfap2c* gene (100%), whereas crRNA design 1 showed lower editing efficiencies (*Gata3*: 51%, *Tfap2c*: 62%). While crRNA design 2 was proven to result in lower frameshift efficiencies for *Gata2* (70%) and *Tfap2a* (0%), crRNA design 1 exhibited more optimal frameshift efficiencies for both genes (*Gata2*: 75%, *Tfap2a*: 46%).

Based on frameshift efficiencies observed in the abovementioned experiment, crRNA designs were selected for each gene to conduct experiments on the embryo level with (**Table 12**).

**Table 12:** Selected crRNA designs per gene.

Gene	Selected crRNA design
<i>Gata3</i>	Design 2
<i>Gata2</i>	Design 1
<i>Tfap2c</i>	Design 2
<i>Tfap2a</i>	Design 1



**Figure 14: On-target efficiencies of CRISPR/Cas9 complexes containing both design 1 and 2 for each gene of interest (GOI).** (A) *Gata3* (B) *Gata2* (C) *Tfp2c* (D) *Tfp2a*. Mouse ESCs were nucleofected with either scrambled or *Gata2*, *Gata3*, *Tfp2a* or *Tfp2c* targeting CRISPR/Cas9 complexes. Following, DNA extraction was performed, the region surrounding the cleavage site was amplified using PCR, and amplicons were analyzed using Illumina™ Miseq NGS, followed by examination of NGS data with BATCH-GE. Editing efficiencies amounting <5% were omitted, considering a 5% error margin.

The most frequently observed mutations caused by the selected crRNA designs for each GOI are displayed in **Table 13**. For the selected crRNA designs for each GOI, mostly caused mutations at the on-target site were all reported to be deletions. It must be highlighted that InDelphi predicts the type of deletions in a very accurate manner, as the selected *Gata3*, *Gata2* and *Tfp2c* crRNA designs result in the predicted deletions (at highly similar relative frequencies). However, the *Tfp2a* crRNA design 1 prediction was inaccurate, as a deletion of 2 bp most frequently occurred at the on-target site, rather than the predicted 17 bp deletion. Yet, it must be noted that a deletion of 17 bp was second most frequently reported at the on-target site, at a frequency of 15.45%. As the deletions at the on-target site for the selected *Gata2*, *Gata3* and *Tfp2c* crRNA designs cause a shift in the ORF, and show a small size, they are desired for our applications.

**Table 13:** Most frequently observed mutations for the selected crRNA designs per gene of interest, accompanied by InDelphi predictions.

Gene	Design	Type of mutation	Sequence	RF (%)	InDelphi prediction
<b>Gata3</b>	Design 2	DEL 1	gccgccacag[t]gggtagagg	14.35	DEL 1 12.6%
<b>Gata2</b>	Design 1	DEL 7	gacacagtag[tggacca]tggaggtggc	33.89	DEL 7 39.5%
<b>Tfap2c</b>	Design 2	DEL 5	gaaagctgcc[cacgt]cactctcctc	26.18	DEL 5 38.7%
<b>Tfap2a</b>	Design 1	DEL 2	accgtgccgt[cc]cgttgctggt	22.47	DEL 17 43.9%

### 3.2.2 Off-target effects

Subsequent to crRNA selection, the presence of off-target effects was examined at the top 10 *in-silico* predicted off-target sites for the *Gata2*, *Gata3*, *Tfap2a* and *Tfap2c* targeting CRISPR/Cas9 complexes (containing the selected crRNA design). **Supplementary tables 10-13** display the sequences of the top 10 *in-silico* predicted off-target sites of *Gata2*, *Gata3*, *Tfap2c* and *Tfap2a* targeting CRISPR/Cas9 complexes, alongside their respective PAM sites, location in the genome, and the number of mismatches compared to the on-target site. First, using Primer3Plus, primers were designed for the top 10 *in-silico* predicted off-target sites for each gene, and tested on naïve mESC DNA (WT). Subsequently, fragment analysis was performed to determine their potency to amplify the off-target sites and their specificity to the respective regions. The size (bp) of the amplicons generated by the primers for the top 10 *in-silico* predicted off-target sites per GOI are shown in **Supplementary figures 2-5**. Upon primer optimization, the regions surrounding the top 10 *in-silico* predicted off-target sites were amplified for each GOI, using DNA extracted from naïve mESCs targeted for either *Gata3*, *Gata2*, *Tfap2c* or *Tfap2a* (see section 3.2.1). Amplicons were further analyzed with the Illumina™ Miseq NGS platform, followed by examination of the obtained NGS data with BATCH-GE to assess for possible off-target effects. **Tables 14-17** display editing efficiencies of the top 10 *in-silico* predicted off-target sites per GOI for CRISPR/Cas9 targeted, scrambled and WT naïve mESCs. As can be observed in **Tables 14-17**, no considerable off-target effects (meaning resulting in a relative editing efficiency > 5%) among the top 10 *in-silico* predicted off-target sites for our GOIs were detected. Even if screening for off-target effects was performed using a lower threshold for editing of 2%, no noteworthy off-target effects were detected. However, it must be acknowledged that analysis failed for certain samples which therefore could not be assessed for off-target editing. For these samples, the same procedure will be repeated in the future.

**Table 14:** Off-target editing efficiencies for *Gata3* after nucleofection with RNP complex containing crRNA design 2, compared to the off-target editing efficiencies observed in scrambled and WT mESC DNA extracts. (\*) = analysis failed, OT= off target site

Off-target	Editing efficiency KO	Editing efficiency scrambled targeted	Editing efficiency WT
OT1	0	0.0068	0
OT2	0.0026	0.0038	0
OT3	0	0	0
OT4	0.0049	0.0026	0.0023
OT5	0.0108	0.0182	0.0044
OT6	0.0011	0.0037	0.0030
OT7	0.0093	0.0031	*
OT8	0	0	*
OT9	0	0	0.0013
OT10	0.0045	0.0064	0

**Table 15:** Off-target editing efficiencies for *Gata2* after nucleofection with RNP complex containing crRNA design 1, compared to the off-target editing efficiencies observed in scrambled and WT mESC DNA extracts. (\*) = analysis failed, OT= off target site

Off-target	Editing efficiency KO	Editing efficiency scrambled targeted	Editing efficiency WT
OT1	*	*	*
OT2	0.0017	0.0026	0
OT3	0	0.0017	0.0007
OT4	0	0.0050	0.0012
OT5	0	0	0
OT6	0.0051	0.0116	0.0019
OT7	0.0093	*	*
OT8	*	*	*
OT9	*	*	0.0052
OT10	0.0061	0	0

**Table 16:** Off-target editing efficiencies for *Tfap2c* after nucleofection with RNP complex containing crRNA design 2, compared to the off-target editing efficiencies observed in scrambled and WT mESC DNA extracts. (\*) = analysis failed, OT= off target site

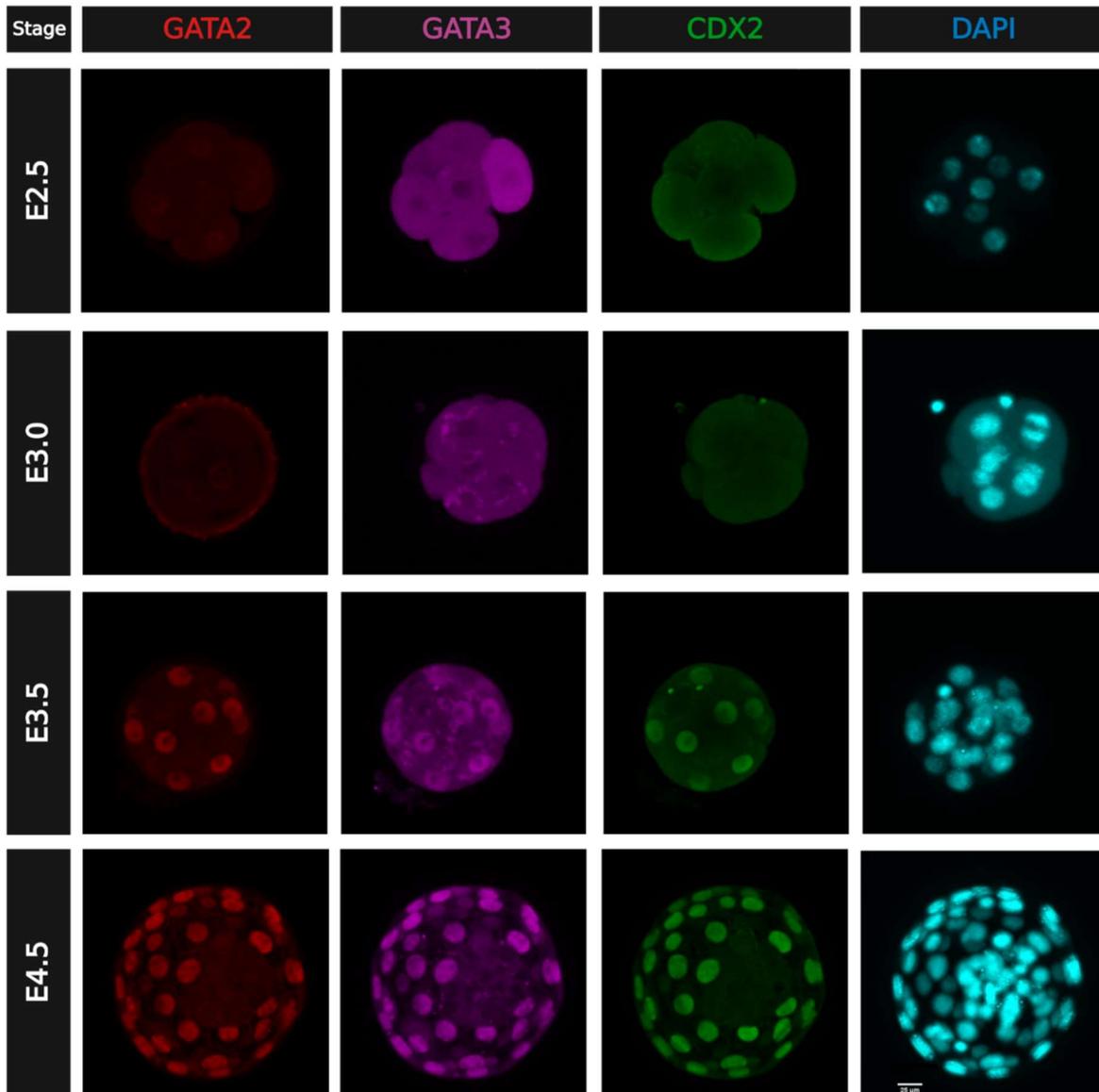
Off-target	Editing efficiency KO	Editing efficiency scrambled targeted	Editing efficiency WT
OT1	0	0.0023	0.0012
OT2	0.0060	0	0.0270
OT3	0.0090	0.107	0.0034
OT4	0.0039	0	0.0012
OT5	0	0	0
OT6	*	*	*
OT7	*	*	0
OT8	0.0022	0.0025	0.0019
OT9	0.005	0	0.0007
OT10	0	0.0063	0.0121

**Table 17:** Off-target editing efficiencies for *Tfap2a* after nucleofection with RNP complex containing crRNA design 1, compared to the off-target editing efficiencies observed in scrambled and WT mESC DNA extracts. (\*) = failed in run, OT= off target site

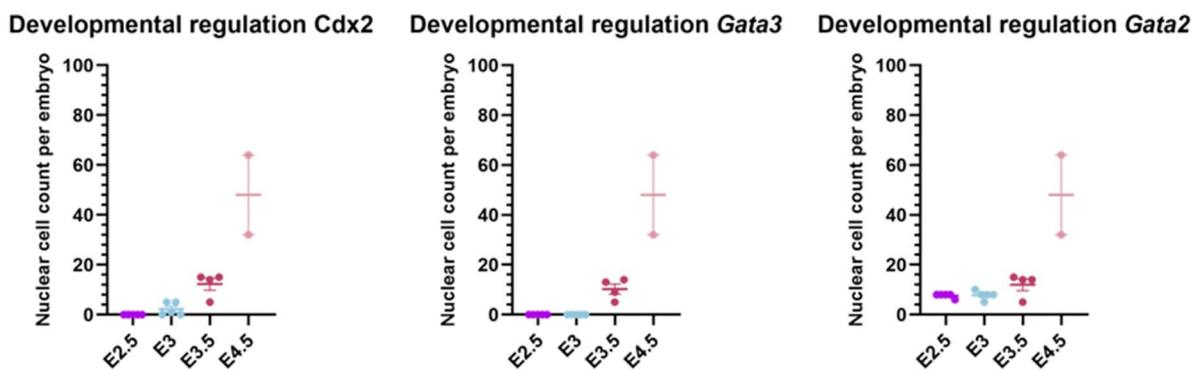
Off-target	Editing efficiency KO	Editing efficiency scrambled targeted	Editing efficiency WT
OT1	0	0	0.0020
OT2	0.0020	0.0018	0.0009
OT3	0.009	0	0.0010
OT4	0	0	0
OT5	0.0042	0	0.0029
OT6	0	0	0
OT7	0	0	0
OT8	0.0109	0	0.0073
OT9	0.0020	0.0021	0.0015
OT10	*	0	0

### 3.3 Immunofluorescent staining

To have baseline information on the expression patterns of GATA2 and GATA3, immunofluorescent staining of GATA3, GATA2 and CDX2 (TE marker) was performed in WT mouse embryos at different stages of embryonic development (E2.5 (n=5), E3.0 (n=5), E3.5 (n=4) and E4.5 (n=2)), and compared to single-cell transcriptomic datasets provided by Blakeley *et al.*<sup>50</sup>. DAPI was used to stain DNA present in the nucleus. The nucleus was considered positive for the stained protein if signal intensity was higher in the nucleus, in comparison to the cytoplasm. It should be mentioned that this staining was executed by Gwenny Cosemans (mentor) in the Department of Stem Cell and Developmental Biology in Leuven, as confocal imaging was performed there. **Figure 15** displays confocal imaging of the immunofluorescent stained embryos, whereas **Figure 16** exhibits quantification of GATA2/3 and CDX2 positive nuclei. As can be observed in **Figure 15**, CDX2 (green) is expressed in the cytoplasm of the blastomeres at E2.5 and E3.0, preceding its restriction to the TE lineage at E3.5 and E4.5. GATA2 (red) exhibits very low nuclear expression at E2.5 and E3.0, whereas expression is co-localized with CDX2 at E3.5 and E4.5, restricting it to the TE lineages at these stages in development. Besides, at E2.5 and E3.0, GATA3 (purple) is present in the cytoplasm, in addition to cytoplasmic aggregation at E3.0, preceding nuclear expression at E3.5 and E4.5. Corresponding with GATA2, at E3.5 and E4.5, GATA3 is expressed in the nuclei of TE cells (CDX2+) and is thus restricted to this lineage. Finally, it should be noted that at both E3.5 and E4.5, GATA2 and GATA3 expressions are co-localized in nuclei of TE cells (CDX2+).



**Figure 15:** Confocal imaging of immunofluorescent staining of mouse embryos for GATA3, GATA2, CDX2 and DAPI at E2.5, E3.0, E3.5 and E4.5 (WT). Images were taken at 20x maximum projection.

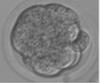
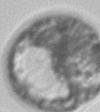
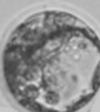
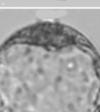
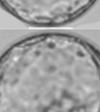


**Figure 16:** Quantification of GATA2, GATA3 and CDX2 positive nuclei, of E2.5, E3.0, E3.5 and E4.5 WT mouse embryos, following immunofluorescent staining of E2.5, E3.0, E3.5 and E4.5 WT mouse embryos. Nuclei were considered positive for the immunofluorescent stained protein if signal intensity in the nucleus was higher in comparison to signal intensity in the cytoplasm.

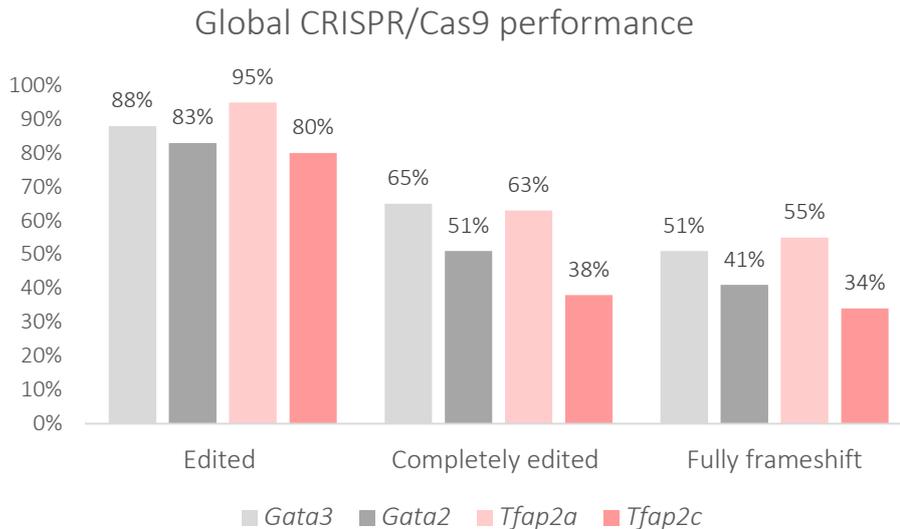
### 3.4 Targeting *Gata3/Gata2* and *Tfap2a/Tfap2c* in mouse embryos via CRISPR/Cas9

Subsequent to the optimization and testing of CRISPR/Cas9 components in mESCs, the *Gata2*, *Gata3*, *Tfap2a* and *Tfap2c* targeting CRISPR/Cas9 RNP complexes were applied in mouse zygotes to produce a KO of our GOIs. Delivery was performed via electroporation, which allows for pore formation in the plasma membrane of the zygote using electrical pulses (**Figure 10**). Moreover, in this context, RNP complexes were used, as there is a lower possibility of off-target effects compared to the use of plasmids for delivery of the CRISPR/Cas9 components, since exposure time of the DNA to the CRISPR/Cas9 is lower. Single embryos were morphologically scored on E4.5 (as displayed in **Table 18**) before DNA extraction and PCR amplification of the region surrounding the target site of either the *Gata3*, *Gata2*, *Tfap2c* or *Tfap2a* targeting CRISPR/Cas9 complex. Subsequently, amplicons were examined with Illumina™ Miseq NGS platform, followed by analysis of NGS data with BATCH-GE to assess on-target editing efficiencies.

**Table 18:** E4.5 grading system.

Score	Morphology	Example
Compacted morula	Compacted embryo with 16-32 cells, no cavitation	
Grade 1 blastocyst	Blastocyst with a cavity smaller than 50% of embryo surface	
Grade 2 blastocyst	Blastocyst with a cavity larger than 50% of embryo surface	
Grade 3 blastocyst	Expanded blastocyst	
Grade 4 blastocyst	Hatching blastocyst	

General on-target performance of the *Gata2*, *Gata3*, *Tfap2a* and *Tfap2c* targeting CRISPR/Cas9 complexes in mouse embryos are displayed in **Figure 17**. **Table 19** shows the absolute number of edited, completely edited or frameshift-only embryos, per GOI, obtained from 13 experimental replicates. It must be acknowledged that only samples with sufficient sequencing depth (50 read pairs) were included. Embryos containing a frameshift efficiency of minimally 95% are considered fully KO (-/-), (taking into account a 5% error margin) and are thus desired for our applications. Analysis of NGS data revealed that 88% (n=43) of *Gata3* targeted embryos (n=49) displayed editing, whereas only 65% (n=32) was completely edited, and 51% (n=25) were reported to contain only frameshift mutations in all cells. In addition, a slightly less successful performance of the *Gata2* and *Tfap2c* targeting CRISPR/Cas9 complexes was reported, as lower editing efficiencies in all three categories were observed. However, high editing efficiencies were found in the *Tfap2a* targeted embryos (n=38), with global editing efficiencies of 95% (n=36), 63% (n=24) of embryos being completely edited, and 55% (n=21) containing only frameshift mutations.

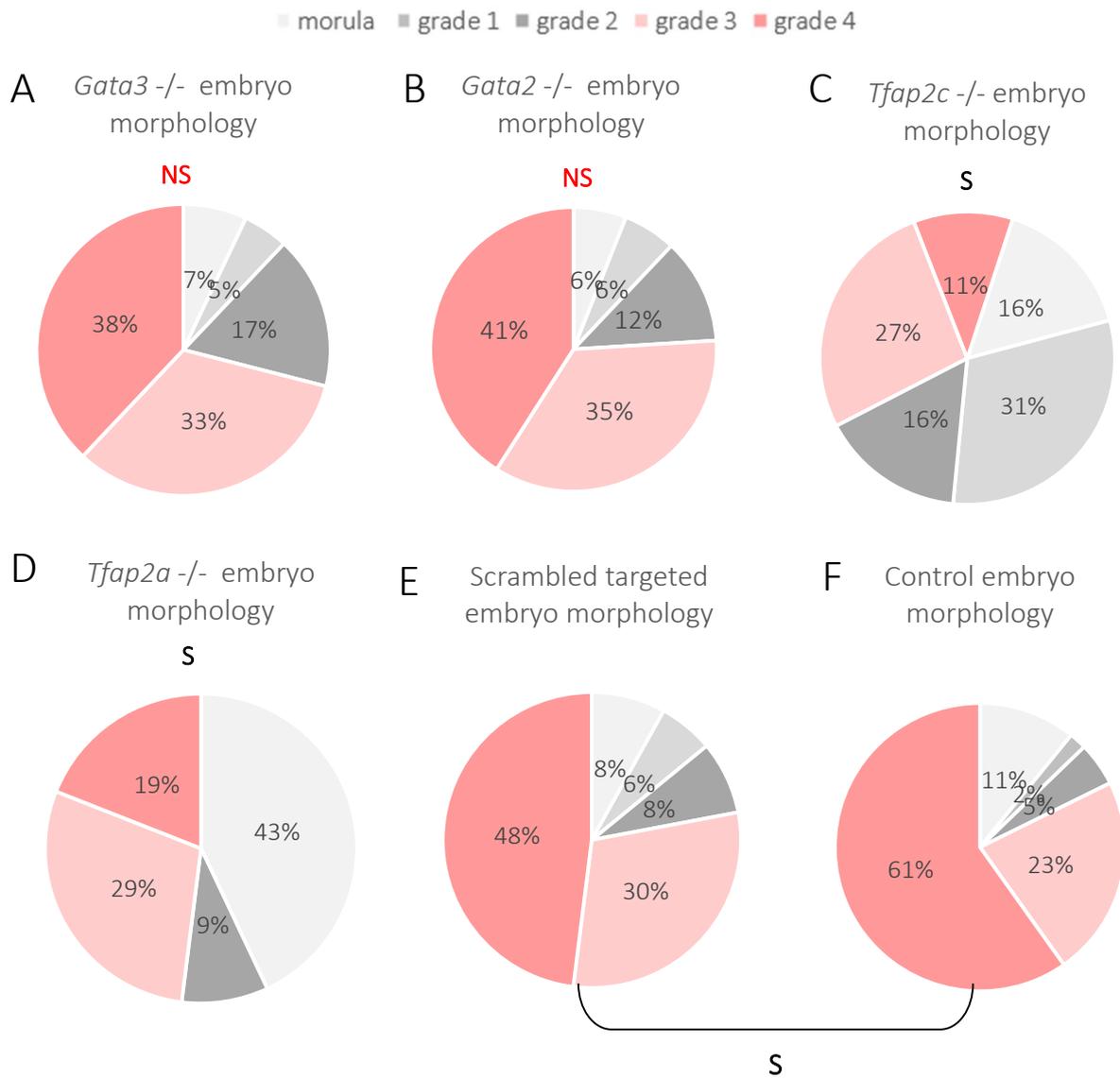


**Figure 17: Global CRISPR/Cas9 performance for each gene of interest.** Mouse embryos were electroporated with *Gata2*, *Gata3*, *Tfap2a* and *Tfap2c* targeting CRISPR/Cas9 complexes and cultured until E4.5. Following, DNA was extracted, the region surrounding the cleavage site was amplified using PCR, and amplicons were analyzed with Illumina™ Miseq NGS. NGS data was examined with BATCH-GE, providing us with on-target editing efficiencies.

**Table 19: Absolute number of embryos which were edited, completely edited or which contained only frameshift mutations, per gene of interest.**

Gene	Total number of embryos	Edited	Fully edited	Frameshift only
<b>Gata3</b>	49	43	32	25
<b>Gata2</b>	41	34	21	17
<b>Tfap2a</b>	38	36	24	21
<b>Tfap2c</b>	56	45	21	19

To further explore the potential morphological effects of the KO of our GOIs, morphology of KO embryos is displayed in **Figure 18**, alongside the morphology of scrambled targeted and control embryos. **Table 20** exhibits the number of *Gata2* *-/-*, *Gata3* *-/-*, *Tfap2a* *-/-*, *Tfap2c* *-/-*, scrambled and control embryos adopting the morula, grade 1, grade 2, grade 3 or grade 4 blastocyst morphology. Typical morphology of *Gata2* *-/-*, *Gata3* *-/-*, *Tfap2a* *-/-* and *Tfap2c* *-/-* embryos is shown in **Supplementary figure 6**. First, to assess the effect of the procedure (electroporation and delivery of CRISPR/Cas9) on embryo morphology, a Chi square test was performed to compare both scrambled targeted (n=96) and control embryo (n=107) morphology. As a statistically significant difference was observed, morphological distributions of the embryos targeted for our GOIs were separately compared to scrambled embryo morphology using a Chi square test, rather than to control embryos. **Supplementary table 14** displays the statistical output for each test. As can be seen in **Figure 18**, the majority of *Gata3* *-/-* (93%, n=22), *Gata2* *-/-* (94%, n=16) and *Tfap2c* *-/-* (84%, n=16) embryos were still able to form blastocysts. However, 57% of *Tfap2a* *-/-* embryos (n=12) could develop to the blastocyst stage, whereas 43% displayed morula arrest. Moreover, no statistically significant difference between morphological distributions of *Gata3* *-/-* (n=24) and *Gata2* *-/-* (n=17) embryos compared to scrambled (n=96) embryos was detected. The morphological distribution of *Tfap2c* *-/-* (n=19) and *Tfap2a* *-/-* (n=21) embryos and scrambled targeted embryos did, however, differ in a statistically significant way, compared to scrambled morphology.



**Figure 18: Morphological distribution of 100% KO embryos for each gene of interest.** In each figure, significance is displayed following comparison of morphologic distribution of gene KO embryos to the phenotypic distribution of scrambled embryos via a Chi Square test. (A) Morphology of 100% *Gata3* *-/-* embryos. (B) Morphology of 100% *Gata2* *-/-* embryos. (C) Morphology of 100% *Tfap2c* *-/-* embryos. (D) Morphology of 100% *Tfap2a* *-/-* embryos. (E) Morphology of scrambled targeted embryos. (F) Morphology of control embryos. NS= non-significant, S= significant

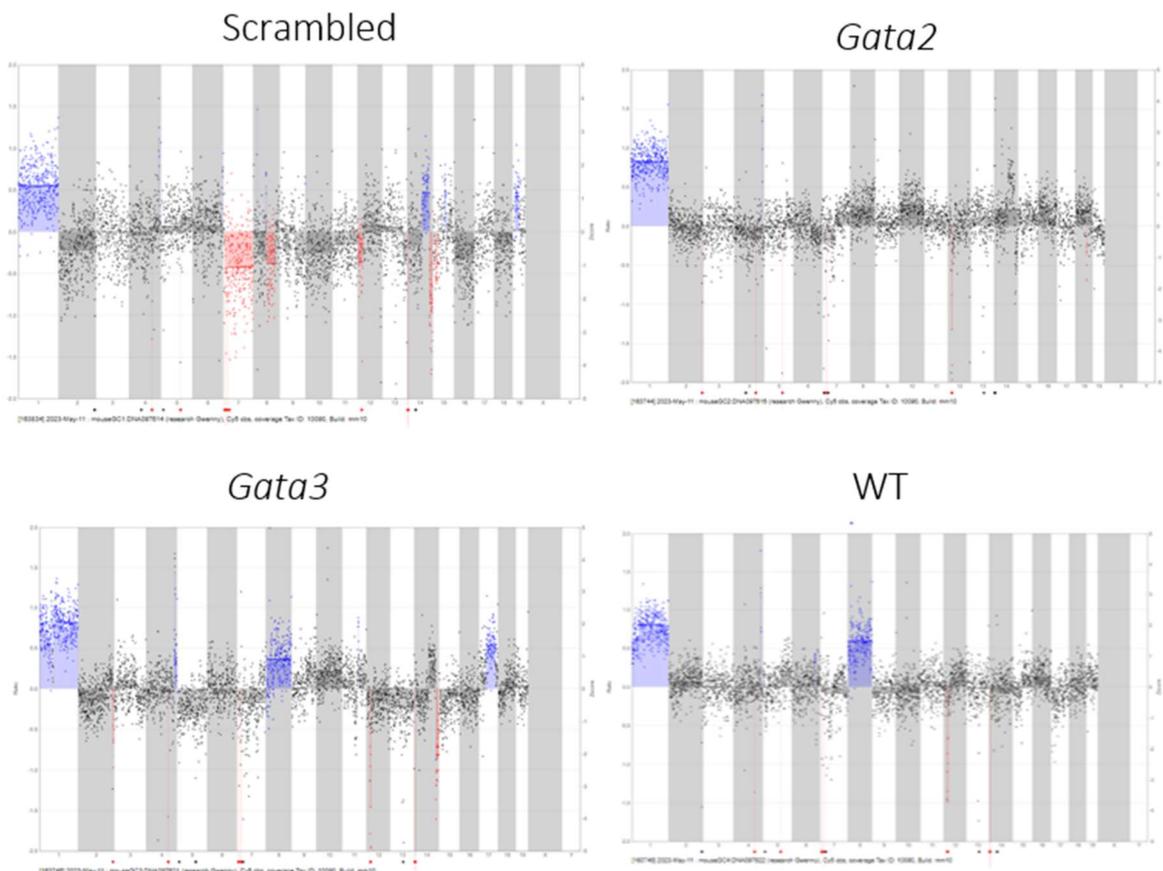
**Table 20: Number of 100% KO embryos adopting the morula, grade 1, grade 2, grade 3 or grade 4 phenotype.**

Gene	Morula	Grade 1	Grade 2	Grade 3	Grade 4	Total
<b><i>Gata3</i></b>	2	1	4	8	9	24
<b><i>Gata2</i></b>	1	1	2	6	7	17
<b><i>Tfap2c</i></b>	3	6	3	5	2	19
<b><i>Tfap2a</i></b>	9	0	2	6	4	21
<b>scrambled</b>	7	6	8	28	46	96
<b>Control</b>	11	2	5	24	65	107

### 3.5 Shallow whole genome sequencing to investigate (partial) chromosome loss

Previously, it was described that genome editing by the CRISPR/Cas9 system might result in significant (partial) chromosome losses or large deletions at the on-target site<sup>84,86</sup>. As the occurrence of these events could cause the phenotype of an embryo rather than the KO of one of our GOIs, these events are not desired. Therefore, the occurrence of large deletions/(partial) chromosome loss was investigated more in depth in naïve mESCs, which was used as a screening tool in this context.

To assess for the potential occurrence of (partial) chromosome loss due to the activity of CRISPR/Cas9, shallow whole genome sequencing (WGS) was performed on DNA from either scrambled targeted mESCs or mESCs targeted with the CRISPR/Cas9 complex containing the selected designs for *Gata3* and *Gata2* (**Figure 19**). This was performed as shallow WGS allows for the detection of aneuploidy and chromosomal imbalances<sup>89</sup>. A trisomy of chromosome 1 was observed in all profiles, whereas an additional trisomy of chromosome 8 was also seen in both the *Gata3* targeted and WT mESCs. A monosomy of chromosome 7, large deletions of chromosomes 8, 12 and 14 and duplication of part of chromosomes 14 and 19 were also found in the scrambled targeted mESCs, in addition to a trisomy of chromosome 17 and a large deletion in chromosome 14 which were present in the *Gata3* targeted mESCs.



**Figure 19:** Shallow WGS profiles generated from scrambled targeted naïve mESCs or naïve mESCs nucleofected with either *Gata3* or *Gata2* targeting CRISPR/Cas9 complexes.

## 4. Discussion

### 4.1 Optimization of the CRISPR/Cas9 components in mouse ESCs

First, using the online in-silico tools Benchling and InDelphi, two crRNAs were designed for each GOI (*Gata3*, *Gata2*, *Tfap2c* and *Tfap2a*). Benchling contains algorithms which score the crRNA designs based on on- or off-target efficiency, while InDelphi, on the other hand, anticipates whether the crRNA design will result in in-frame or out-of-frame mutations, and displays the anticipated on-target occurring indels, as well as the microhomology strength of the cleavage site.

Prior to applying the CRISPR/Cas9 components in mouse embryos, their activity was evaluated in naïve mESCs, since they exhibit translational value for CRISPR/Cas9 editing in mouse embryos<sup>47</sup>. Mouse ESCs can display either naïve or primed pluripotency. While primed mESCs appear to be similar to the more differentiated mouse EPI, naïve mESCs are considered the *in vitro* equivalent of the early ICM, and present a less differentiated state of pluripotency. In comparison to primed mESCs, naïve mESCs display a number of advantages, including faster, easier and scalable expansion, in addition to better survival after single-cell passaging and less spontaneous differentiation<sup>101</sup>. Therefore, naïve mESCs were used for the optimization of CRISPR/Cas9 components since they present more advantageous characteristics. In addition, since our GOIs are not expressed in naïve mESCs, this results in minimal consequences of gene KO in naïve mESCs, making this screening tool optimal. More specifically, the R1WT naïve mESC stem cell line was used. To test the activity of the CRISPR/Cas9 components in naïve mESCs, RNP complexes were first prepared and delivered to the mESCs via nucleofection, a technique used to transfect cells using electrical pulses. DNA was extracted from the mESCs, and regions of interest (either on- or off-target sites) were amplified via PCR. Subsequently, amplicons were analyzed with the Illumina™ Miseq NGS platform, followed by analysis of the NGS data with BATCH-GE, to assess for editing efficiency. Besides, for the selected crRNA designs per GOI, most frequently occurring mutations at the on-target site were evaluated. It must be highlighted that it was proven that highly accurate predictions were made by InDelphi, regarding the type (insertion/deletion) and size (bp) of the most commonly occurring on-target mutation since the chosen crRNA designs for *Gata3* (DEL1), *Gata2* (DEL7) and *Tfap2c* (DEL5) produced the expected deletions, at remarkably similar relative frequencies. InDelphi thus proves to be extremely valuable for predicting the nature of indels occurring at the on-target site for a particular crRNA design. In contrast, the prediction was inaccurate for the *Tfap2a* targeting crRNA, as a 2 bp deletion occurred at the on-target site, whereas a 17 bp deletion was foreseen. However, it must be acknowledged that a 17 bp deletion was observed to be second most frequently occurring at the on-target site, at a relative frequency of 15.45%. In that regard, InDelphi typically demonstrates excellent translational value, yet not consistently. Considering these results, in the future, human naïve mESCs will also be used as a screening tool before applying CRISPR/Cas9 components in human embryos, taking into account that different crRNA designs will have to be designed for human applications.

Subsequently, considering both broad editing and frameshift efficiencies, one out of two crRNA designs for each GOI was chosen. The highest frameshift efficiencies were demonstrated by crRNA design 2 for both *Gata3* and *Tfap2c*, and crRNA design 1 for the *Gata2* and *Tfap2a*. These designs were therefore selected for further use during embryo experiments. It must be highlighted that the impact of CRISPR/Cas9 editing is not entirely

comparable in R1 naïve mESCs and mouse embryos. Consequently, mESCs are only used as a screening method to increase the likelihood of obtaining 100% KO embryos, as mouse embryos are valuable material, and ethical constraints are considered.

Upon selection of one crRNA design per GOI, the potential occurrence of off-target effects was examined at the top 10 *in-silico* predicted off-target sites per gene. No significant off-target effects (defined as editing efficiencies >5%) were present for the CRISPR/Cas9 complexes targeting *Gata2*, *Gata3*, *Tfap2a* and *Tfap2c*. Nevertheless, off-target effects due to CRISPR/Cas9 activity should be noted with caution, as such effects may result in the phenotype of the embryo, rather than the KO of the gene of interest. In the future, if significant off-target effects were to be observed, a crRNA targeting the same domain can be designed, considering other off-target effects are present for this crRNA, to confirm that the phenotype of the KO embryo is caused due to on-target editing, rather than off-target editing. If the same phenotype were to be observed with both crRNA designs, we can conclude the phenotype is a result of the gene KO, rather than off-target effects. In addition, it is important to acknowledge that off-target editing was evaluated using DNA extracted from a whole well of naïve mESCs (containing countless colonies) rather than single colonies. For this reason, there is a chance that minor off-target editing was concealed, since a low number of off-target editing events would be difficult to detect among large number of unedited off-target sites. Hence, DNA extraction from CRISPR/Cas9 targeted single colonies will be carried out in the future, followed by assessment for off-target effects at the top 10 *in-silico* predicted off-target sites for each gene, to examine this. Moreover, it must be pointed out that only the top 10 *in-silico* predicted off-target sites were tested for off-target effects. By comparing the number of mismatches between the off-target site and the on-target region, Benchling predicts a ranking among off-target sites in descending order, from most likely to obtain off-target editing, to least likely, in an exponential way. Even though off-target effects can occur genome wide (on locations besides these top 10 *in-silico* predicted off-target sites), it is very unlikely that off-target effects would take place at locations containing more mismatches than our top 10 *in-silico* predicted off-target sites. However, to exclude this possibility, deep whole genome sequencing of CRISPR/Cas9 targeted naïve mESCs could be performed, which would allow us to check for off-target effects genome wide.

#### 4.2 Immunofluorescent staining of GATA2/3 at different embryonic stages

At various embryonic stages (E2.5, E3.0, E3.5 and E4.5), immunofluorescent staining of GATA2, GATA3 and CDX2 was carried out in WT embryos, to examine their expression levels. CDX2 is widely known as a nuclear marker for TE cells and was therefore used to assess the nuclear expression of GATA2 and GATA3 in TE cells. The cell's nucleus was considered positive for a certain protein if signal intensity was higher compared to signal intensity in the cytoplasm. Although single-cell transcriptomic data from mouse preimplantation embryos (originating from research performed by Blakeley *et al.*<sup>50</sup>) already revealed the expression pattern of GATA2/3, we wanted to confirm this using immunofluorescent staining. Nuclear CDX2 expression was detected at both E2.5 and at E3.5, preceding nuclear expression restricted to TE cells at E3.5 and E4.5. Relying on the single-cell transcriptomic data from Blakeley *et al.*, it was expected that *Gata2* expression would only be detectable from the morula stage onwards, followed by increased expression levels in the mouse blastocyst<sup>50</sup>. In contrast to the findings of Blakeley *et al.*<sup>50</sup>, upon confocal visualization of mouse embryos stained for GATA2, low levels of nuclear GATA2 expression were reported at E2.5 and E3.0, followed by restriction of GATA2 expression to

the nuclei of TE cells (CDX2+), at E3.5 and E4.5. Likewise, when observing immunofluorescent staining performed by Home *et al.*, nuclear GATA2 expression was also exhibited at these early stages, in addition to a continuous nuclear expression in TE cells, up until the blastocyst stage, which is in line with our findings<sup>69</sup>. Furthermore, from the E2.5 stage onwards, continuous expression of GATA3 is detected. As the single-cell transcriptomic data from Blakeley *et al.* revealed detectable expression of GATA3 from the 4-cell stage onwards, cytoplasmic expression of GATA3 at these early stages was an expected result<sup>50</sup>. However, it must be mentioned that localization of GATA3 expression differs among developmental stages of the preimplantation embryo. At both E2.5 and E3.0, GATA3 is expressed in the cytoplasm of the blastomeres, whereas at E3.5 and E4.5, GATA3 expression becomes restricted to TE cells, as overlapping expression with CDX2+ nuclei are observed. This is in line with the findings of Home *et al.*, as restriction of GATA3 expression to the TE cells of the blastocyst was also reported<sup>68</sup>. Interestingly, at E3.0 and E3.5, GATA3 exhibits cytoplasmic aggregation in the different blastomeres. Yet, the reason behind this cytoplasmic aggregation remains unelucidated and needs extensive research in the future, since such cytoplasmic aggregation has not been observed before. Additionally, it must be emphasized that even though GATA2 and CDX2 expressions are co-localized in every embryonic stage, there is not always co-localization of GATA3 and CDX2 expressions. Besides, co-localization between CDX2 and GATA3 expression only occurs at the E3.5 and E4.5 stages, whereas this is not the case at E2.5 and E3.0. Interestingly, it must be indicated that nuclear GATA2 expression precedes GATA3 expression, which could indicate that GATA2 transcriptional activity precedes the expression of GATA3. In conclusion, the onset of GATA2 expression is CDX2 independent, since CDX2 nuclear expression of GATA2 and CDX2 overlap completely. As opposed to GATA2, GATA3 expression is potentially dependent on CDX2 (and GATA2), as nuclear CDX2 (and GATA2) expression preceded restriction of GATA3 to the TE nuclei.

### 4.3 Targeting *Gata3*, *Gata2*, *Tfap2c* and *Tfap2a* with CRISPR/Cas9 in mouse embryos

After CRISPR/Cas9 components were refined and tested in naïve mESCs, they were delivered to mouse zygotes to produce either *Gata3*<sup>-/-</sup>, *Gata2*<sup>-/-</sup>, *Tfap2a*<sup>-/-</sup> or *Tfap2c*<sup>-/-</sup> embryos. Embryos were then cultured up to E4.5, and scored for morphology (**Table 18**), prior to DNA extraction. Following, the regions surrounding the on-target site were genotyped as described above.

#### 4.3.1 On-target editing efficiency

First, overall performance of the *Gata2*, *Gata3*, *Tfap2a* and *Tfap2c* targeting CRISPR/Cas9 complexes in mouse embryos was evaluated, by examining the proportion of edited, completely edited and frameshift-only containing embryos. In comparison to *Gata3* (51%), *Gata2* (41%) and *Tfap2c* targeted embryos (34%), *Tfap2a* targeted embryos show the highest percentage of embryos containing only frameshift mutations (55%). Since only part of the edited embryos exhibit editing in all cells (for all GOIs), genetic mosaicism still took place in some of the embryos and can be caused by a multitude of reasons. First, half-life of CRISPR/Cas9 RNP complexes is +/- 24 hours, therefore, CRISPR/Cas9 remains active following first cell division. This way, editing could potentially have occurred following first

cleavage division of the zygote, leading to certain cells being edited whereas other cells remained unedited. It is also possible that upon DSB generation by CRISPR/Cas9 prior to the first cell division, NHEJ restored the DSB correctly (meaning without the creation of small insertions or deletions). Therefore, it could potentially have occurred that DSB generation prior to the first cell division did not lead to the formation of indels at the on-target site, but the CRISPR/Cas9 complex was able to bind the same site again (as no mutations were inserted) and caused frameshift mutations only after the first division. Yet, it still must be noted that our method serves its purpose in avoiding mosaicism in the biggest part of the completely edited embryos. Besides, when investigating embryos targeted for our GOI, it was observed that only part of the embryos which were fully edited, contained only frameshift mutations. It must also be pointed out that a variety of frameshift mutations occur at the same on-target site, which may result in variation within the *Gata3*<sup>-/-</sup>, *Gata2*<sup>-/-</sup>, *Tfap2c*<sup>-/-</sup> and *Tfap2a*<sup>-/-</sup> embryo groups. Unfortunately, the types of mutations that take place at the on-target site are largely out of our control. However, when creating the crRNAs, only those designs which are predicted (by InDelphi) to cause frameshift mutations, are chosen.

Overall, it can be concluded from the results displayed in **Figure 17** that on-target editing efficiency via high-throughput electroporation-based methods is highly gene specific. This can be due to a multitude of reasons. First, differing potency of the crRNA designs could play a role in this, whereas the location of the gene might contribute to the accessibility of the DNA for the CRISPR/Cas9 complex. Besides, due to the formation of euchromatin or heterochromatin at specific stages of embryonic development, certain gene locations could be more amenable for CRISPR/Cas9 targeting. Interestingly, a strong correlation between GC content of the gRNA and on-target editing efficiency has been reported in the literature<sup>102-104</sup>, which is something to keep into account when designing crRNAs in the future.

#### 4.3.2 Morphology of *Gata3*<sup>-/-</sup>, *Gata2*<sup>-/-</sup>, *Tfap2c*<sup>-/-</sup> and *Tfap2a*<sup>-/-</sup> mouse embryos

First, it should be pointed out that following a Chi Square test, a statistically significant difference was observed between the morphological distributions of scrambled targeted and control embryos (which did not undergo the electroporation procedure) following a Chi square test. This finding suggests that the used methods (electroporation, CRISPR/Cas9 delivery) have an impact on embryo morphology. Therefore, morphological distributions of KO embryos for our GOIs were separately compared to scrambled embryo morphology, rather than control embryo morphology, by making use of a Chi square test.

Based on findings of Home *et al.* it was predicted that *Gata3*<sup>-/-</sup> embryos would not be able to develop into blastocysts, since their results suggested that downregulation of GATA3 prevented morula to blastocyst transformation<sup>68</sup>. Additionally, in a follow-up study of Home *et al.*, a conditional KO of *Gata3* was created in mouse embryos, which revealed partially impaired formation of blastocysts<sup>69</sup>. Nevertheless, in our study, 93% of *Gata3*<sup>-/-</sup> embryos were able to form blastocysts (5% grade 1, 17% grade 2, 33% grade 3, 38% grade 4) whereas only 7% experienced morula arrest. Also, a Chi square statistical analysis revealed no significant difference between the morphological distribution of *Gata3*<sup>-/-</sup> and scrambled targeted embryos. Yet, it must be pointed out that a decreased blastocyst quality can be observed for *Gata3*<sup>-/-</sup> embryos (38% grade 4, 33% grade 3, 17% grade 2, 5% grade 1), in comparison to scrambled targeted embryos (48% grade 4, 30% grade 3, 8% grade 2, 6% grade 1). Based on aforementioned findings, it was hypothesized that, in case the *Gata3* mRNA is not degraded by the NMD pathway, GATA3 activity might originate from a still functional (untargeted) TAD domain, since this domain is located before the targeted ZnF

domain. First, an immunofluorescent staining will be performed with an antibody binding at a location N-terminal from the DNA binding domain. If signal for GATA3 is still detected, a protein product is still formed and the *Gata3* gene should be targeted earlier in the coding sequence. If this were to be observed in the future, two crRNAs will be designed targeting the TAD domain, and will be used to target the *Gata3* gene in the same way as described above. Likewise, a comparison can be made between *Gata3* *-/-* embryos targeted for the DNA binding ZnF domain and *Gata3* *-/-* embryos targeted for the TAD domain. In contrast, it is still possible that *Gata3* is not essential for TE formation, and, consequently, for blastocyst formation. This theory is supported by the study of Pandolfi *et al.*, since *Gata3* *-/-* embryos were reported to be able to form blastocysts<sup>105</sup>. Interestingly, *Gata3* *-/-* embryos only showed lethality between E11-E12 (post-implantation stage), which therefore lead to the conclusion that *Gata3* function is not necessary for blastocyst formation, yet it is for implantation<sup>105</sup>. However, our findings on embryo morphology of *Gata3* *-/-* embryos still must be supported by future experiments and no conclusions can be made based on the abovementioned results. Besides, the paper of Home *et al.* suggested that GATA2 was not necessary for the blastocyst formation, as *Gata2* *-/-* embryos could develop into blastocysts and no significant morula arrest was reported<sup>69</sup>. Hence, based on these results, it was expected that *Gata2* *-/-* embryos would still be able to form blastocysts. The findings from Home *et al.* were supported by abovementioned results, as 94% of *Gata2* *-/-* embryos were able to reach the blastocyst stage (6% grade 1, 12% grade 2, 35% grade 3, 41% grade 4), whereas only 6% of *Gata2* *-/-* embryos experienced morula arrest. Additionally, no statistically significant difference between the morphological distribution of *Gata2* *-/-* and scrambled targeted embryos was observed following the performance of a Chi square statistical analysis. Furthermore, it was suggested by Ma *et al.* that there could be functional redundancy between *Gata2* and *Gata3*, which could result in the ability of both *Gata3* *-/-* and *Gata2* *-/-* embryos to reach the blastocyst stage<sup>106</sup>. Yet, to be able to make this conclusion, *Gata2* *-/-* *Gata3* *-/-* DKO embryos should be generated, and phenotype of these embryos should be assessed. Other than that, *Gata2/3* *-/-* DKO models should also be submitted to downstream protein analysis, in addition to transcriptomic studies on the RNA level, to potentially conclude functional redundancy. Besides, extensive genetic characterization of these DKO models is necessary, since the chance of the occurrence of chromosomal aberrations/large deletions due to CRISPR/Cas9 activity increases as two regions are targeted within the same embryo.

The morphological distribution of scrambled targeted compared to *Tfap2a* *-/-* and *Tfap2c* *-/-* embryo morphology, did, however, differ in a statistically significant way. As can be observed in **Figure 18**, 57% of *Tfap2a* *-/-* embryos were able to reach the blastocyst stage, while 43% experienced morula arrest. Given the high rate of morula arrest, these results point to a potential role of *Tfap2a* in TE formation. However, more research is required to draw this conclusion. In contrast, only 16% of *Tfap2c* *-/-* embryos exhibited morula arrest, whereas 84% of *Tfap2c* *-/-* embryos were still able to develop into blastocysts. However, a higher proportion of *Tfap2c* *-/-* embryos exhibiting morula arrest was anticipated, as a recent article of Zhu *et al.* demonstrated the crucial function of *Tfap2c* in polarization of mouse embryos<sup>8</sup>. It was hypothesized that if polarization cannot occur due to *Tfap2c* *-/-*, TE formation (and thus blastocyst formation) would be disturbed. Yet still 84% of *Tfap2c* *-/-* embryos were able to form blastocysts.

#### 4.4 Shallow whole genome sequencing of WT, scrambled, *Gata2* and *Gata3* targeted naïve mESC

It is widely known that (partial) chromosome loss can occur due to CRISPR/Cas9 activity<sup>84,86</sup>. To investigate the occurrence of CRISPR/Cas9 induced (partial) chromosome loss, shallow WGS of either scrambled, WT, *Gata2* or *Gata3* targeted mESCs was performed.

Several chromosomal aberrations can be seen in the shallow WGS profiles from both scrambled targeted and untargeted WT mESCs. First, in the shallow WGS profile of scrambled targeted mESCs, a trisomy of chromosome 1 is exhibited, alongside a monosomy of chromosome 7, large deletions in chromosomes 8,12 and 14 and duplications in certain regions of chromosomes 14 and 19. Besides, a trisomy of chromosome 1 in addition to a trisomy of chromosome 8 is observed in WT mESCs. High passaging numbers (P40-50) may have contributed to the chromosomal aberrations in the WT mESCs, since each passaging increases the likelihood of aneuploidy. Interestingly, a trisomy of chromosome 8 has been reported to provide mESCs with a growth advantage, which may account for why WT, scrambled and *Gata3* targeted mESCs all exhibited this trisomy<sup>107</sup>. In contrast to WT mESCs, in the scrambled targeted mESCs, several other chromosomal aberrations are present (monosomy of chromosome 7, partial chromosome loss of chromosome 8, large deletions of chromosomes 12 and 14 and duplications of certain regions in chromosomes 14 and 19). These chromosomal aberrations might have been caused by Cas9 activity or the nucleofection procedure itself, as they are not present in WT mESCs. However, since the crRNA in scrambled CRISPR/Cas9 complexes lack a homologous site in the DNA, it is unlikely that the complex will bind randomly some place in the DNA. In the future, this experiment will be repeated to exclude this possibility.

Furthermore, shallow WGS of *Gata3* targeted mESCs revealed a trisomy of chromosomes 1, 8 and 17, in addition to a large deletion at chromosome 14. A trisomy of chromosomes 1 and 8 and a large deletion of chromosome 14 is also present in the WT mESCs and could therefore be due to the intrinsic genomic composition (resulting from high passaging number). However, a trisomy of chromosome 17 is not present in the scrambled targeted mESC DNA, nor in the WT sample, and could therefore be caused by high passaging number or the nucleofection technique. Given recent reports that CRISPR/Cas9 editing might result in significant duplications, it cannot be ruled out that this was caused by CRISPR/Cas9 activity<sup>108</sup>. However, the murine *Gata3* gene is located on chromosome 2, and none of its off-target sites are located at chromosome 17. As a result, it is unlikely that off-target CRISPR/Cas9 activity caused the trisomy of chromosome 17, but this possibility can, however, not be excluded.

In addition, shallow WGS of the *Gata2* targeted mESCs displayed a trisomy of chromosome 1, which is also present in the WT, scrambled and *Gata3* targeted mESCs. Hence, this is most likely due to high passaging number. Furthermore, no other chromosomal aberrations were reported, which indicates that no chromosome loss is caused by the *Gata2* targeting CRISPR/Cas9 complex. It should be mentioned that only one sample was tested, and that we cannot rule out the occurrence of large deletions caused by *Gata2* targeting CRISPR/Cas9 complexes in all future samples.

In conclusion, the occurrence of large deletions and (partial) chromosome loss due to CRISPR/Cas9 activity has to be further investigated in both embryos and mESCs, as it could not be ruled out that certain chromosomal aberrations were caused by CRISPR/Ca9 activity.

## 5. Conclusion

First, two crRNA designs (designed using Benchling and InDelphi), were applied in naïve mESCs to test their activity. For the *Gata3* and *Tfap2c* gene, crRNA design 2 was the most optimal for our applications, whereas for *Gata2* and *Tfap2a* crRNA design 1 was selected, based on broad editing and frameshift efficiencies. Subsequent to analysis of the top 10 *in-silico* predicted off-target sites in mESCs targeted for either *Gata3*, *Gata2*, *Tfap2c* or *Tfap2a*, no notable off-target effects were found for all of our GOIs (meaning with a higher editing efficiency than 5%). Continuously, CRISPR/Cas9 editing of *Gata3*, *Gata2*, *Tfap2a* and *Tfap2c* in mouse embryos revealed high global editing efficiencies, yet a notable smaller percentage of embryos only displayed frameshift mutations, which is desired for our applications. It must be acknowledged that when comparing the phenotype of scrambled targeted and control embryos using a Chi square test, a statistically significant difference was observed. Therefore, morphology of *Gata2* *-/-*, *Gata3* *-/-*, *Tfap2a* *-/-* and *Tfap2c* *-/-* embryos was directly compared to the morphological distribution of Scrambled targeted embryos, rather than control embryo morphology. Furthermore *Gata3* *-/-*, *Gata2* *-/-*, *Tfap2a* *-/-* and *Tfap2c* *-/-* embryos were still able to form blastocysts, yet a significantly differing embryo phenotype was reported when comparing the morphological distributions of *Tfap2a* *-/-* and *Tfap2c* *-/-* embryos to scrambled targeted embryo morphology. These results report a first indication of the role of *Tfap2a* and *Tfap2c* in blastocyst formation, and more specifically in formation of the TE. Additionally, immunofluorescent staining of GATA2/3 in WT embryos revealed that very low nuclear GATA2 expression is observed at the E2.5 and E3.0 stage, whereas GATA2 expression becomes restricted to the TE nuclei (CDX2+) at the blastocyst stage (E3.5 and E4.5). Moreover, GATA3 is expressed in mouse embryos beginning at the E2.5 stage up until the blastocyst stage. However, whereas GATA3 exhibits expression in the cytoplasm of the blastomeres of both E2.5 and E3.0 embryos, from the E3.5 stage onwards, GATA3 expression becomes restricted to the nucleus of the TE cells (CDX2+). Interestingly, GATA3 also shows cytoplasmic aggregation at the E2.5 and E3.0 stage, of which the meaning has not been elucidated to date. Chromosomal aberrations present in both the WT, scrambled and *Gata2/3* targeted mESCs are most likely caused due to high passaging number, whereas aneuploidies detected in the scrambled mESCs (and also in *Gata2* and *Gata3* targeted mESCs) could be a result of aspecific Cas9 activity or of the nucleofection technique. Lastly, the trisomy of chromosome 17 in the *Gata3* targeted mESCs are most likely caused due to high passaging number or the nucleofection technique, yet it cannot be ruled out that this resulted from CRISPR/Cas9 activity. In the future, the consequences of *Gata2* *-/-*, *Gata3* *-/-*, *Tfap2a* *-/-* and *Tfap2c* *-/-* will be examined on both the protein and RNA level. By performing immunofluorescent staining of a predetermined set of proteins, the molecular consequences of either *Gata3* *-/-*, *Gata2* *-/-*, *Tfap2c* *-/-* and *Tfap2a* *-/-* can be investigated. Additionally, the transcriptomic consequences of the KO of our GOIs on the blastocyst level will be examined using RNA sequencing. This can be performed on both whole embryos (blastocysts) and at the single-cell level. The latter will be carried out to specifically look into the transcriptomic changes separately in the TE and the ICM cells. Besides, the combinatorial loss of *Gata2/3* or *Tfap2a/c* will also be studied by creating DKO models. Also, the function of GATA2, GATA3, TFAP2A and TFAP2C will be researched in human embryos, to possibly reveal previously unknown interspecies differences.

## 6. References

- 1 Adamson, G. D. *et al.* International Committee for Monitoring Assisted Reproductive Technology: world report on assisted reproductive technology, 2011. *Fertility and sterility* **110**, 1067-1080 (2018).
- 2 Petraglia, F., Serour, G. I. & Chapron, C. The changing prevalence of infertility. *International Journal of Gynecology & Obstetrics* **123**, S4-S8 (2013).
- 3 Fasouliotis, S. J. & Schenker, J. G. Failures in assisted reproductive technology: an overview. *European Journal of Obstetrics and Gynecology and Reproductive Biology* **107**, 4-18 (2003).
- 4 Molè, M. A., Weberling, A. & Zernicka-Goetz, M. Comparative analysis of human and mouse development: From zygote to pre-gastrulation. *Current topics in developmental biology* **136**, 113-138 (2020).
- 5 Chazaud, C. & Yamanaka, Y. Lineage specification in the mouse preimplantation embryo. *Development* **143**, 1063-1074 (2016).
- 6 Niakan, K. K., Han, J., Pedersen, R. A., Simon, C. & Pera, R. A. R. Human pre-implantation embryo development. *Development* **139**, 829-841 (2012).
- 7 White, M., Bissiere, S., Alvarez, Y. D. & Plachta, N. Mouse embryo compaction. *Current topics in developmental biology* **120**, 235-258 (2016).
- 8 Zhu, M. *et al.* Developmental clock and mechanism of de novo polarization of the mouse embryo. *Science* **370**, eabd2703 (2020).
- 9 Zhu, M. *et al.* Human embryo polarization requires PLC signaling to mediate trophoblast specification. *Elife* **10**, e65068 (2021).
- 10 Malassiné, A. & Cronier, L. Hormones and human trophoblast differentiation. *Endocrine* **19**, 3-11 (2002).
- 11 Strumpf, D. *et al.* Cdx2 is required for correct cell fate specification and differentiation of trophoblast in the mouse blastocyst. (2005).
- 12 Tanaka, S., Kunath, T., Hadjantonakis, A.-K., Nagy, A. & Rossant, J. Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282**, 2072-2075 (1998).
- 13 Chang, C.-W., Wakeland, A. K. & Parast, M. M. Trophoblast lineage specification, differentiation and their regulation by oxygen tension. *Journal of Endocrinology* **236**, R43-R56 (2018).
- 14 James, J., Carter, A. M. & Chamley, L. Human placentation from nidation to 5 weeks of gestation. Part I: What do we know about formative placental development following implantation? *Placenta* **33**, 327-334 (2012).
- 15 Louvet, S., Aghion, J., Santa-Maria, A., Mangeat, P. & Maro, B. Ezrin becomes restricted to outer cells following asymmetrical division in the preimplantation mouse embryo. *Developmental biology* **177**, 568-579 (1996).
- 16 Plusa, B. *et al.* Downregulation of Par3 and aPKC function directs cells towards the ICM in the preimplantation mouse embryo. *Journal of cell science* **118**, 505-515 (2005).
- 17 Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. Dynamics of global gene expression changes during mouse preimplantation development. *Developmental cell* **6**, 117-131 (2004).
- 18 Hirate, Y. *et al.* Polarity-dependent distribution of angiominin localizes Hippo signaling in preimplantation embryos. *Current biology* **23**, 1181-1194 (2013).
- 19 Sasaki, H. in *Seminars in cell & developmental biology*. 80-87 (Elsevier).
- 20 Wicklow, E. *et al.* HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst. *PLoS genetics* **10**, e1004618 (2014).
- 21 Keramari, M. *et al.* Sox2 is essential for formation of trophoblast in the preimplantation embryo. *PLoS one* **5**, e13952 (2010).
- 22 Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642 (2003).
- 23 Palmieri, S. L., Peter, W., Hess, H. & Schöler, H. R. Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Developmental biology* **166**, 259-267 (1994).
- 24 Karasek, C., Ashry, M., Driscoll, C. S. & Knott, J. G. A tale of two cell-fates: role of the Hippo signaling pathway and transcription factors in early lineage formation in mouse preimplantation embryos. *Molecular human reproduction* **26**, 653-664 (2020).
- 25 Niwa, H., Miyazaki, J.-i. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature genetics* **24**, 372-376 (2000).

- 26 Nishioka, N. *et al.* The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Developmental cell* **16**, 398-410 (2009).
- 27 Mihajlović, A. I. & Bruce, A. W. Rho-associated protein kinase regulates subcellular localisation of Angiomotin and Hippo-signalling during preimplantation mouse embryo development. *Reproductive BioMedicine Online* **33**, 381-390 (2016).
- 28 Shi, X. *et al.* Rho differentially regulates the Hippo pathway by modulating the interaction between Amot and Nf2 in the blastocyst. *Development* **144**, 3957-3967 (2017).
- 29 Graham, S. J. *et al.* BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. *Nature communications* **5**, 1-11 (2014).
- 30 Home, P. *et al.* Altered subcellular localization of transcription factor TEAD4 regulates first mammalian cell lineage commitment. *Proceedings of the National Academy of Sciences* **109**, 7362-7367 (2012).
- 31 De Paepe, C. *et al.* BMP4 plays a role in apoptosis during human preimplantation development. *Molecular reproduction and development* **86**, 53-62 (2019).
- 32 Oron, E. & Ivanova, N. Cell fate regulation in early mammalian development. *Physical biology* **9**, 045002 (2012).
- 33 Petropoulos, S. *et al.* Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012-1026 (2016).
- 34 Niakan, K. K. & Eggan, K. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Developmental biology* **375**, 54-64 (2013).
- 35 Kuijk, E. W. *et al.* The roles of FGF and MAP kinase signaling in the segregation of the epiblast and hypoblast cell lineages in bovine and human embryos. *Development* **139**, 871-882 (2012).
- 36 Christodoulou, N. *et al.* Morphogenesis of extra-embryonic tissues directs the remodelling of the mouse embryo at implantation. *Nature communications* **10**, 1-12 (2019).
- 37 Mobley, R. J. *et al.* MAP3K4 controls the chromatin modifier HDAC6 during trophoblast stem cell epithelial-to-mesenchymal transition. *Cell reports* **18**, 2387-2400 (2017).
- 38 Wang, J., Mayernik, L. & Armant, D. R. Trophoblast adhesion of the peri-implantation mouse blastocyst is regulated by integrin signaling that targets phospholipase C. *Developmental biology* **302**, 143-153 (2007).
- 39 Hertig, A. T. & Rock, J. On the development of the early human ovum, with special reference to the trophoblast of the previllous stage: A description of 7 normal and 5 pathologic human ova. *American Journal of Obstetrics and Gynecology* **47**, 149-184 (1944).
- 40 Lindenberg, S., Hyttel, P., Sjøgren, A. & Greve, T. A comparative study of attachment of human, bovine and mouse blastocysts to uterine epithelial monolayer. *Human Reproduction* **4**, 446-456 (1989).
- 41 Aplin, J. D. & Ruane, P. T. Embryo–epithelium interactions during implantation at a glance. *Journal of cell science* **130**, 15-22 (2017).
- 42 Weimar, C. H., Uiterweer, E. D. P., Teklenburg, G., Heijnen, C. J. & Macklon, N. S. In-vitro model systems for the study of human embryo–endometrium interactions. *Reproductive biomedicine online* **27**, 461-476 (2013).
- 43 Lv, B. *et al.* Single-cell RNA sequencing reveals regulatory mechanism for trophoblast cell-fate divergence in human peri-implantation conceptuses. *PLoS biology* **17**, e3000187 (2019).
- 44 Johnson, M. & Maro, B. The distribution of cytoplasmic actin in mouse 8-cell blastomeres. (1984).
- 45 Nikas, G., Ao, A., Winston, R. M. & Handyside, A. H. Compaction and surface polarity in the human embryo in vitro. *Biology of reproduction* **55**, 32-37 (1996).
- 46 Stamatiadis, P. *et al.* TEAD4 regulates trophectoderm differentiation upstream of CDX2 in a GATA3-independent manner in the human preimplantation embryo. *Human Reproduction* **37**, 1760-1773 (2022).
- 47 Fogarty, N. M. *et al.* Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* **550**, 67-73 (2017).
- 48 Stamatiadis, P. *et al.* Comparative analysis of mouse and human preimplantation development following POU5F1 CRISPR/Cas9 targeting reveals interspecies differences. *Human Reproduction* **36**, 1242-1252 (2021).
- 49 Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715-724 (2010).

- 50 Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3151-3165 (2015).
- 51 Hertig, A. T., Rock, J. & Adams, E. C. A description of 34 human ova within the first 17 days of development. *American Journal of Anatomy* **98**, 435-493 (1956).
- 52 Christodoulou, N. *et al.* Sequential formation and resolution of multiple rosettes drive embryo remodelling after implantation. *Nature cell biology* **20**, 1278-1289 (2018).
- 53 Eckert, D., Buhl, S., Weber, S., Jäger, R. & Schorle, H. The AP-2 family of transcription factors. *Genome biology* **6**, 1-8 (2005).
- 54 Li, H., Sheridan, R. & Williams, T. Analysis of TFAP2A mutations in branchio-oculo-facial syndrome indicates functional complexity within the AP-2 $\alpha$  DNA-binding domain. *Human molecular genetics* **22**, 3195-3206 (2013).
- 55 Meier, P., Koedood, M., Philipp, J., Fontana, A. & Mitchell, P. J. Alternative mRNAs encode multiple isoforms of transcription factor AP-2 during murine embryogenesis. *Developmental biology* **169**, 1-14 (1995).
- 56 Ohtaka-Maruyama, C., Hanaoka, F. & Chepelinsky, A. B. A novel alternative spliced variant of the transcription factor AP2 $\alpha$  is expressed in the murine ocular lens. *Developmental biology* **202**, 125-135 (1998).
- 57 Bauer, R. *et al.* The genomic structure of the human AP-2 transcription factor. *Nucleic acids research* **22**, 1413-1420 (1994).
- 58 Hilger-Eversheim, K., Moser, M., Schorle, H. & Buettner, R. Regulatory roles of AP-2 transcription factors in vertebrate development, apoptosis and cell-cycle control. *Gene* **260**, 1-12 (2000).
- 59 Bai, Q. *et al.* Dissecting the first transcriptional divergence during human embryonic development. *Stem Cell Reviews and Reports* **8**, 150-162 (2012).
- 60 Kuckenbergh, P., Kubaczka, C. & Schorle, H. The role of transcription factor Tcfap2c/TFAP2C in trophoctoderm development. *Reproductive biomedicine online* **25**, 12-20 (2012).
- 61 Winger, Q., Huang, J., Auman, H. J., Lewandoski, M. & Williams, T. Analysis of transcription factor AP-2 expression and function during mouse preimplantation development. *Biology of reproduction* **75**, 324-333 (2006).
- 62 Richardson, B. D., Cheng, Y.-H., Langland, R. A. & Handwerger, S. Differential expression of AP-2 $\gamma$  and AP-2 $\alpha$  during human trophoblast differentiation. *Life sciences* **69**, 2157-2165 (2001).
- 63 Viger, R. S., Guittot, S. M., Anttonen, M., Wilson, D. B. & Heikinheimo, M. Role of the GATA family of transcription factors in endocrine development, function, and disease. *Molecular endocrinology* **22**, 781-798 (2008).
- 64 Bates, D. L., Chen, Y., Kim, G., Guo, L. & Chen, L. Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA. *Journal of molecular biology* **381**, 1292-1306 (2008).
- 65 Zhou, Y. *et al.* Dynamic regulation of GATA2 in fate determination in hematopoiesis: possible approach to hPSC-derived hematopoietic stem/progenitor cells. *Blood Science* **2**, 1-6 (2020).
- 66 Adomas, A. B. *et al.* Breast tumor specific mutation in GATA3 affects physiological mechanisms regulating transcription factor turnover. *BMC cancer* **14**, 1-14 (2014).
- 67 Cosemans, G. *et al.* CRISPR/Cas9 mediated knock-out (KO) reveals a divergent role for trophoctoderm markers GATA2/3 in the mouse and human preimplantation embryo. *HUMAN REPRODUCTION* **37** (2022).
- 68 Home, P. *et al.* GATA3 is selectively expressed in the trophoctoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. *Journal of Biological Chemistry* **284**, 28729-28737 (2009).
- 69 Home, P. *et al.* Genetic redundancy of GATA factors in the extraembryonic trophoblast lineage ensures the progression of preimplantation and postimplantation mammalian development. *Development* **144**, 876-888 (2017).
- 70 Bai, H., Sakurai, T., Godkin, J. D. & Imakawa, K. Expression and potential role of GATA factors in trophoblast development. *Journal of Reproduction and Development* **59**, 1-6 (2013).
- 71 Krendl, C. *et al.* GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophoctoderm with repression of pluripotency. *Proceedings of the National Academy of Sciences* **114**, E9579-E9588 (2017).
- 72 Ma, Y., Zhang, L. & Huang, X. Genome modification by CRISPR/Cas9. *The FEBS journal* **281**, 5186-5193 (2014).
- 73 Redman, M., King, A., Watson, C. & King, D. What is CRISPR/Cas9? *Archives of Disease in Childhood-Education and Practice* **101**, 213-215 (2016).

- 74 Bekaert, B. *et al.* in *Seminars in Cell & Developmental Biology*. (Elsevier).
- 75 Yang, H. *et al.* Methods favoring homology-directed repair choice in response to  
CRISPR/Cas9 induced-double strand breaks. *International journal of molecular sciences* **21**,  
6461 (2020).
- 76 Jang, H.-K., Song, B., Hwang, G.-H. & Bae, S. Current trends in gene recovery mediated by  
the CRISPR-Cas system. *Experimental & Molecular Medicine* **52**, 1016-1027 (2020).
- 77 Tuladhar, R. *et al.* CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA  
misregulation. *Nature communications* **10**, 4056 (2019).
- 78 Dumitrache, L. C. *et al.* Trex2 enables spontaneous sister chromatid exchanges without  
facilitating DNA double-strand break repair. *Genetics* **188**, 787-797 (2011).
- 79 Liu, M. *et al.* Methodologies for improving HDR efficiency. *Frontiers in genetics* **9**, 691 (2019).
- 80 Ata, H. *et al.* Robust activation of microhomology-mediated end joining for precision gene  
editing applications. *PLoS genetics* **14**, e1007652 (2018).
- 81 Manghwar, H. *et al.* CRISPR/Cas systems in genome editing: methodologies and tools for  
sgRNA design, off-target evaluation, and strategies to mitigate off-target effects. *Advanced  
science* **7**, 1902312 (2020).
- 82 Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing  
in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research* **24**, 1012-  
1019 (2014).
- 83 Lamas-Toranzo, I. *et al.* Strategies to reduce genetic mosaicism following CRISPR-mediated  
genome edition in bovine embryos. *Scientific reports* **9**, 1-8 (2019).
- 84 Adikusuma, F. *et al.* Large deletions induced by Cas9 cleavage. *Nature* **560**, E8-E9 (2018).
- 85 Egli, D. *et al.* Inter-homologue repair in fertilized human eggs? *Nature* **560**, E5-E7 (2018).
- 86 Lee, H. & Kim, J.-S. Unexpected CRISPR on-target effects. *Nature biotechnology* **36**, 703-704  
(2018).
- 87 Alanis-Lobato, G. *et al.* Frequent loss of heterozygosity in CRISPR-Cas9–edited early human  
embryos. *Proceedings of the National Academy of Sciences* **118**, e2004832117 (2021).
- 88 Togi, S., Ura, H. & Niida, Y. Optimization and Validation of Multimodular, Long-Range PCR–  
Based Next-Generation Sequencing Assays for Comprehensive Detection of Mutation in  
Tuberous Sclerosis Complex. *The Journal of Molecular Diagnostics* **23**, 424-446 (2021).
- 89 Deleye, L. *et al.* Shallow whole genome sequencing is well suited for the detection of  
chromosomal aberrations in human blastocysts. *Fertility and sterility* **104**, 1276-1285. e1271  
(2015).
- 90 Scholefield, J. & Harrison, P. T. Prime editing—an update on the field. *Gene Therapy* **28**, 396-  
401 (2021).
- 91 Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or  
donor DNA. *Nature* **576**, 149-157 (2019).
- 92 Porto, E. M., Komor, A. C., Slaymaker, I. M. & Yeo, G. W. Base editing: advances and  
therapeutic opportunities. *Nature Reviews Drug Discovery* **19**, 839-859 (2020).
- 93 Gaudelli, N. M. *et al.* Programmable base editing of A• T to G• C in genomic DNA without DNA  
cleavage. *Nature* **551**, 464-471 (2017).
- 94 Habib, O., Habib, G., Hwang, G.-H. & Bae, S. Comprehensive analysis of prime editing  
outcomes in human embryonic stem cells. *Nucleic acids research* **50**, 1187-1197 (2022).
- 95 Liang, P. *et al.* Correction of  $\beta$ -thalassemia mutant by base editor in human embryos. *Protein  
& cell* **8**, 811-822 (2017).
- 96 Liang, P. *et al.* Effective gene editing by high-fidelity base editor 2 in mouse zygotes. *Protein &  
cell* **8**, 601-611 (2017).
- 97 Schene, I. F. *et al.* Prime editing for functional repair in patient-derived disease models.  
*Nature communications* **11**, 5352 (2020).
- 98 Mou, H. *et al.* CRISPR/Cas9-mediated genome editing induces exon skipping by alternative  
splicing or exon deletion. *Genome biology* **18**, 1-8 (2017).
- 99 Boel, A. *et al.* BATCH-GE: Batch analysis of Next-Generation Sequencing data for genome  
editing assessment. *Scientific reports* **6**, 1-10 (2016).
- 100 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature biotechnology* **29**, 24-26 (2011).
- 101 Bai, Q. *et al.* Temporal analysis of genome alterations induced by single-cell passaging in  
human embryonic stem cells. *Stem cells and development* **24**, 653-662 (2015).
- 102 Hall, B. *et al.* Genome editing in mice using CRISPR/Cas9 technology. *Current Protocols in  
Cell Biology* **81**, e57 (2018).
- 103 Ren, F. *et al.* Efficiency optimization of CRISPR/Cas9-mediated targeted mutagenesis in  
grape. *Frontiers in plant science* **10**, 612 (2019).

- 104 Ren, X. *et al.* Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized  
sgRNA parameters in *Drosophila*. *Cell reports* **9**, 1151-1162 (2014).
- 105 Pandolfi, P. P. *et al.* Targeted disruption of the GATA3 gene causes severe abnormalities in  
the nervous system and in fetal liver haematopoiesis. *Nature genetics* **11**, 40-44 (1995).
- 106 Ma, G. T. *et al.* GATA-2 and GATA-3 regulate trophoblast-specific gene expression in vivo.  
*Development* **124**, 907-914 (1997).
- 107 Gaztelumendi, N. & Nogués, C. Chromosome instability in mouse embryonic stem cells.  
*Scientific Reports* **4**, 1-8 (2014).
- 108 Li, J. *et al.* Efficient inversions and duplications of mammalian regulatory DNA elements and  
gene clusters by CRISPR/Cas9. *Journal of molecular cell biology* **7**, 284-298 (2015).
- 109 *Tfap2a*. (2023, February). Retrieved from Ensembl:  
[https://www.ensembl.org/Mus\\_musculus/Gene/Summary?db=core;g=ENSMUSG00000021359;r=13:40868778-40891852](https://www.ensembl.org/Mus_musculus/Gene/Summary?db=core;g=ENSMUSG00000021359;r=13:40868778-40891852)
- 110 *TFAP2A*. (2023, February). Retrieved from Ensembl:  
[https://www.ensembl.org/Homo\\_sapiens/Gene/Summary?db=core;g=ENSG00000137203;r=6:10393186-10419659](https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000137203;r=6:10393186-10419659)
- 111 *Tfap2c*. (2023, February). Retrieved from Ensembl:  
[https://www.ensembl.org/Mus\\_musculus/Gene/Summary?db=core;g=ENSMUSG00000028640;r=2:172391513-172400542](https://www.ensembl.org/Mus_musculus/Gene/Summary?db=core;g=ENSMUSG00000028640;r=2:172391513-172400542)
- 112 *TFAP2C*. (2023, February). Retrieved from Ensembl:  
[https://www.ensembl.org/Homo\\_sapiens/Gene/Summary?db=core;g=ENSG00000087510;r=20:56629306-56639283](https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000087510;r=20:56629306-56639283)
- 113 *Gata2*. (2023, February). Retrieved from Ensembl:  
[https://www.ensembl.org/Mus\\_musculus/Gene/Summary?db=core;g=ENSMUSG00000015053;r=6:88170873-88184014](https://www.ensembl.org/Mus_musculus/Gene/Summary?db=core;g=ENSMUSG00000015053;r=6:88170873-88184014)
- 114 *GATA2*. (2023, February). Retrieved from Ensembl:  
[https://www.ensembl.org/Homo\\_sapiens/Gene/Summary?db=core;g=ENSG00000179348;r=3:128479427-128493201](https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000179348;r=3:128479427-128493201)
- 115 *Gata3*. (2023, February). Retrieved from Ensembl:  
[https://www.ensembl.org/Mus\\_musculus/Gene/Summary?db=core;g=ENSMUSG00000015619;r=2:9861889-9894845](https://www.ensembl.org/Mus_musculus/Gene/Summary?db=core;g=ENSMUSG00000015619;r=2:9861889-9894845)
- 116 *GATA3*. (2023, February). Retrieved from Ensembl:  
[https://www.ensembl.org/Homo\\_sapiens/Gene/Summary?db=core;g=ENSG00000107485;r=10:8045378-8075198](https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000107485;r=10:8045378-8075198)
- 117 *What is fragment analysis?* (n.d.). Retrieved from Thermo Fisher Scientific:  
<https://www.thermofisher.com/be/en/home/life-science/sequencing/sequencing-learning-center/capillary-electrophoresis-information/what-is-fragment-analysis.html#:~:text=Fragment%20analysis%20is%20a%20genetic,comparison%20to%20an%20internal%20standard.>
- 119 *Integrative Genomics Viewer*. Retrieved from <https://software.broadinstitute.org/software/igv/>

All figures were created using [www.biorender.com](http://www.biorender.com).

# Unravelling the divergent functions of major players in early embryonic development via CRISPR/Cas9 gene editing: from mouse to human

Declercq S., Cosemans G., Boel A., Heindryckx B.

Ghent Fertility And Stem cell Team (G-FaST), Department for Reproductive Medicine, Ghent, Belgium



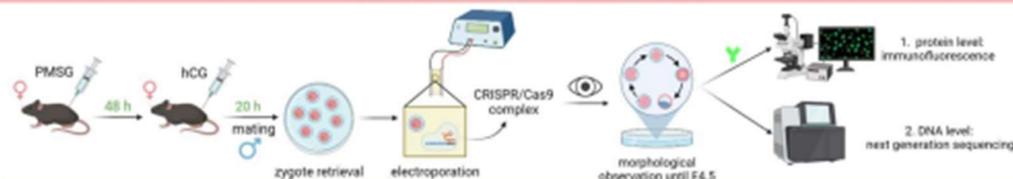
## INTRODUCTION

The recent advancements in clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein9 (Cas9) (**CRISPR/Cas9**) **gene editing** allows researchers to investigate the function of genes which are important during preimplantation development. CRISPR/Cas9 introduces a double stranded break in the target gene, which the cell could repair via an error-prone pathway (non-homologous end-joining). Ultimately, repair could result in the incorporation of insertions or deletions, potentially giving rise to a frameshift mutation, eventually disrupting protein function. We aim to unravel the **molecular mechanisms** involved in trophoblast (TE) formation, the lineage responsible for embryo implantation and placentation. More specifically, we aim to investigate the function of **Gata3** and its isoform **Gata2**, both important TE markers, and thus important to establish implantation of the mouse embryo. Extensive research has been conducted in mouse, yet extrapolation from mouse to human is not always possible [1]. Therefore, in the future, we also aim to investigate our target genes directly in human to unravel possible **interspecies differences**.

## OBJECTIVES

- Successfully create **gene KO** of **Gata3** and its isoform **Gata2** in mouse embryos by using **CRISPR/Cas9**.
- Investigate the function of transcription factors **Gata3**, and its isoform **Gata2** in **polarization, first and second lineage segregation** via the use of **CRISPR/Cas9** gene editing.
- Long term: Unravelling previously unknown **interspecies differences** between molecular mechanisms in **mouse** and **human** TE formation.

## MATERIALS AND METHODS



## RESULTS

### 1. gRNA design

For both the murine **Gata3** and **Gata2** genes, the DNA binding domain was targeted as this would result in defective DNA binding of the transcription factors. Figure 1 shows which exons were targeted in both the mouse **Gata3** and **Gata2** gene.

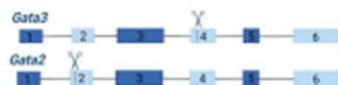


Figure 1: investigated genes and their targeted exons

### 2. Optimization electroporation

To obtain 100% **Gata3/Gata2** KO in mouse embryos, we tested three set-ups with differing electroporation parameters. Parameter values for each set-up are displayed in Table 1.

Set-up	Poring pulse					Transfer pulse					Efficiency
	#P	V	PD (ms)	PI (ms)	D+P	#P	V	PD (ms)	PI (ms)	D+P	
1	4	40	3.5	50	10%	5	5	50	50	40% +/-	
2	6	30	3	100	10%	5	5	50	50	40% +/-	
3	4	40	3	50	10%	5	7	50	50	40% +/-	

Table 1: parameter values for each electroporation set-up. #P= number of pulses, V=voltage, PD= pulse duration, PI= pulse interval, D+P= decay + polarity

### 3. On-target efficiency

Following DNA extraction of the electroporated embryos on E4.5, PCR amplification of the genomic region surrounding the target site was performed.

Subsequently, amplicons were analyzed with next generation sequencing to determine on-target editing efficiency (Figure 2-3).

#### parameter efficiency for Gata3

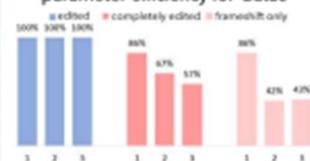


Figure 2: parameter efficiency for Gata3 (one replicate)

#### parameter efficiency for Gata2

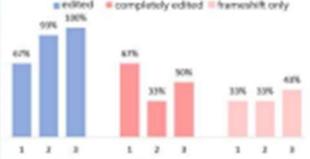


Figure 3: parameter efficiency for Gata2 (one replicate)

### 3. Morphological observation

Following electroporation, 100% KO embryos were scored based on morphology at E4.5 and compared with Scrambled targeted and control embryo morphology (Figure 4).



Figure 4: morphology at E4.5 for Gata3 KO, scrambled and control embryos

As observed in Figure 6, **Gata3**, **Gata2** KO embryos could still form blastocysts. However, we observed affected blastocyst morphology as **Gata3** KO embryos displayed grade 1 blastocyst morphology, whereas **Gata2** KO embryos rather resembled Scrambled embryo morphology.

### 4. Immunofluorescent staining of GATA3 targeted embryos

Immunofluorescent staining was performed on E3.5 **Gata3** KO and wild-type (WT) embryos to confirm the absence of **GATA3** and to assess the effect on the expression of other lineage markers, such as **GATA2** and **NANOG** (Figure 5).

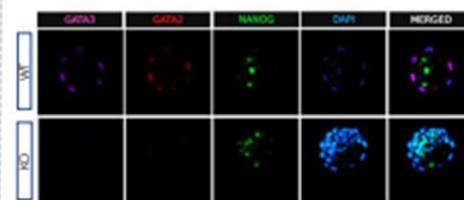


Figure 5: immunofluorescent staining of GATA3, GATA2 and NANOG in a full KO and WT embryo.

## CONCLUSIONS

- **Efficient generation of 100% gene KO** of **Gata3** and **Gata2** genes in mouse embryos could be reached. **Gata3** and **Gata2** KO embryos could still form blastocysts, yet blastocyst morphology was affected (**Gata3** KO embryos were more affected).
- **Scrambled** targeted embryos also showed impaired blastocyst morphology, so in terms of electroporation, a **balance** should be found between editing efficiency and embryo morphology.
- Efficiency via electroporation-based methods is **gene specific**.

[1] Molk, M. A., Weibeling, A. & Zentgraf-Goetz, M. Comparative analysis of human and mouse development: From zygote to pre-gastrulation. *Current topics in developmental biology* 124, 113-138 (2020).

All images created with Biorender.com

## Supplementary

### Supplementary tables

**Supplementary table 1:** Transcripts *Tfap2a* mouse gene. **bold**=functionally most important transcript, source: [www.ensembl.com](http://www.ensembl.com) <sup>(109)</sup>

<i>Tfap2a</i> transcript mouse biotype	Number of exons	Number of amino acids
Protein coding	6	466
Protein coding	<b>7</b>	<b>431</b>
Protein coding	7	437
Protein coding	8	439
Protein coding	7	433
Protein coding	7	439
Protein coding CDS not defined	7	No protein
Protein coding CDS not defined	3	No protein
Retained intron	5	No protein

**Supplementary table 2:** Transcripts *TFAP2A* human gene. **bold** = functionally most important transcript, source: [www.ensembl.com](http://www.ensembl.com) <sup>(110)</sup>

<i>TFAP2A</i> transcript human biotype	Number of exons	Number of amino acids
Protein coding	<b>7</b>	<b>439</b>
Protein coding	<b>7</b>	<b>431</b>
Protein coding	8	439
Protein coding	<b>7</b>	<b>433</b>
Protein coding	<b>6</b>	<b>328</b>
Protein coding	5	218
Protein coding	2	173
Protein coding	3	81
Protein coding	4	136
Nonsense mediated decay	8	113
Nonsense mediated decay	8	121
Protein coding CDS not defined	8	No protein
Protein coding CDS not defined	4	No protein
Protein coding CDS not defined	2	No protein
Protein coding CDS not defined	5	No protein
Protein coding CDS not defined	3	No protein
Retained intron	3	No protein
Retained intron	2	No protein
Retained intron	5	No protein

**Supplementary table 3:** Transcripts *Tfap2c* mouse gene. **bold** = functionally most important transcript, source: [www.ensembl.com](http://www.ensembl.com) <sup>(111)</sup>

<i>Tfap2c</i> transcript mouse biotype	Number of exons	Number of amino acids
Protein coding	7	513
Protein coding	<b>7</b>	<b>449</b>
Protein coding	7	488
Protein coding CDS not known	7	No protein
Protein coding CDS not known	7	No protein
Protein coding CDS not known	2	No protein

**Supplementary table 4:** Transcripts *TFAP2C* human gene. **bold** = functionally most important transcript, source: [www.ensembl.com](http://www.ensembl.com) <sup>(112)</sup>

<i>TFAP2C</i> transcript human biotype	Number of exons	Number of amino acids
Protein coding	<b>7</b>	<b>450</b>
Protein coding	2	118

**Supplementary table 5:** Transcripts *Gata2* mouse gene. **bold** = functionally most important transcript, source: [www.ensembl.com](http://www.ensembl.com) <sup>(113)</sup>

<i>Gata2</i> transcript mouse biotype	Number of exons	Number of amino acids
Protein coding	<b>6</b>	<b>480</b>
Protein coding	<b>6</b>	<b>480</b>
Protein coding	3	34
Retained intron	2	No protein
Retained intron	1	No protein

**Supplementary table 6:** Transcripts *GATA2* human gene. **bold** = functionally most important transcript, source: [www.ensembl.com](http://www.ensembl.com) <sup>(114)</sup>

<i>GATA2</i> transcript human biotype	Number of exons	Number of amino acids
Protein coding	<b>6</b>	<b>480</b>
Protein coding	6	466
Protein coding	<b>7</b>	<b>480</b>
Retained intron	2	No protein
Protein coding	3	150
Protein coding	3	8
Protein coding	8	574
Protein coding	2	76
Protein coding	3	76
Protein coding	5	76
Protein coding CDS not defined	2	No protein
Protein coding	4	76
Protein coding	2	192

**Supplementary table 7:** Transcripts *Gata3* human gene. **bold** = functionally most important transcript, source: [www.ensembl.com](http://www.ensembl.com) <sup>(115)</sup>

<i>Gata3</i> transcript mouse biotype	Number of exons	Number of amino acids
Protein coding	<b>6</b>	<b>443</b>
Protein coding	3	119
Protein coding CDS not defined	2	No protein
Protein coding CDS not defined	2	No protein
Protein coding CDS not defined	3	No protein
Protein coding CDS not defined	2	No protein

**Supplementary table 8:** Transcripts *GATA3* human gene. **bold** = functionally most important transcript, source: [www.ensembl.com](http://www.ensembl.com) <sup>(116)</sup>

<i>GATA3</i> transcript human biotype	Number of exons	Number of amino acids
Protein coding	<b>6</b>	<b>443</b>
Protein coding	6	443
Protein coding	3	191
Protein coding	3	143
Protein coding	2	80

**Supplementary Table 9:** Forward and reverse primer sequences for the target site of both crRNA designs per gene.

Gene	Design	Primer pair
<i>Gata3</i>	Design 1	Forward: CCCTAAACCCTCCTTTTTGC Reverse: TCCCTGGCTGCAAAATCTAC
	Design 2	Forward: CCTCAGCAAGTGGGACATGT Reverse: GCCAGTTTCCTGCAGAAAGC
	TAD1	Forward: CGAGGACATGGAGGTGACTG Reverse: AAGAACTCTCCCAACCAGCG
	TAD2	Forward: TTGCTCACCTTTGCTTCCCA Reverse: CCGAGTTTCCGTAGTAGGGC
<i>Gata2</i>	Design 1	Forward: CGTGTGTGTGACCTCGTGTA Reverse: CTGCGAGTCGAGATGGTTGA
	Design 2	Forward: TTCGCTGAGTTGTGATCCTG Reverse: AGACCACGGACCCCTTTTAC
<i>Tfap2c</i>	Design 1	Forward: GTTTCGCTGTAGCAGGGTTTC Reverse: TCTATGGGGTGAGCCATCTC
	Design 2	Forward: CCCAGTGAGGTCTTCTGCTC Reverse: GCTCACGTCATGGGAAAGTT
<i>Tfap2a</i>	Design 1	Forward: TGGAAGGCCTGCTTCACTT Reverse: GGTGTGTGTGTAAGAGCCCA
	Design 2	Forward: CACACCCTAATGCCGACTT Reverse: CTGAAGGGTGACAATCGCCT

**Supplementary table 10:** Top 10 in-silico predicted off-target sequences of *Gata3* crRNA design 2 by Benchling. red= mismatches

Sequence	PAM	Score	Cut locus	Location		
				Exonic	Intronic	
					Intragenic	Intergenic
GTGCAACCTCT A <b>AC</b> CCCACTG	GGG	3.6	chr1:+1337198 80		<i>Atp2b4</i>	
AGG <b>CCACAT</b> TCT ACCCCACTG	CAG	2.8	chr7:-6987635			x
TGGCA <b>GCT</b> TCT ACCCCACTG	AGG	1.7	chr12:- 7753272		<i>Gm38282</i>	
GG <b>ACAACAC</b> CT ACCCCACTG	GAG	1.6	chr8:- 102945758			X
GG <b>AG</b> AACCTCT <b>G</b> CCCACTG	TGG	1.4	chr12:- 88476116			X
T <b>GTCTAT</b> CTCT ACCCCACTG	GGG	1.0	chr13:+178310 50			x
<b>CCC</b> CA <b>G</b> CTCT ACCCCACTG	TGG	1.0	chr8:+1043447 33		<i>Cmtm3</i>	
GGG <b>AAAG</b> CTCT <b>G</b> CCCACTG	GAG	1.0	chr6:- 122837626		<i>Foxj2</i>	
<b>CAGTA</b> AC <b>CC</b> CT ACCCCACTG	AAG	0.9	chr7:- 43798169		<i>Klk8</i>	
TGG <b>GT</b> AC <b>CTCC</b> ACCCCACTG	TGG	0.9	chr2:- 120037009		<i>Pla2g4b</i>	

**Supplementary table 11:** Top 10 in-silico predicted off-target sequences of *Gata2* crRNA design 1 by Benchling. red=mismatches

Sequence	PAM	Score	Cut locus	Location		
				Exonic	Intronic	
					Intragenic	Intergenic
GACACAG <b>A</b> AGT GGAC <b>C</b> CTGG	GAG	3.1	chr5:+12067 8668	<i>Rasa1</i>		
<b>C</b> ACTCAG <b>A</b> AGT GGACCATGG	TAG	2.8	chr2:- 130591421	<i>Ubox5</i>		
<b>G</b> TACAG <b>C</b> AGT GGACCAT <b>G</b> A	GGG	1.9	chr18:- 66682917			X
<b>G</b> TACAGTAGT GGAC <b>C</b> CTGG	TAG	1.5	chr1:+69097 385			X
<b>G</b> <b>C</b> <b>A</b> AGTAGT GGACCATGG	CAG	1.3	chr4:- 95422059			X
<b>A</b> ACAT <b>T</b> GGT GGACCATGG	AAG	1.2	chr7:- 135000704			X
GAC <b>A</b> AGTAGT GGACCATGG	GAG	1.0	chr4:- 19927914			X
<b>A</b> CCACAG <b>G</b> GG TGGACCATGG	AAG	1.0	chr2:- 174320012			X
GACAC <b>C</b> TATT GGACCATGG	TGG	1.0	chr10:- 98599996			X
<b>A</b> ACCAG <b>G</b> AG <b>A</b> GGACCATGG	GAG	0.9	chr15:+3242 0712			X

**Supplementary table 12:** Top 10 in-silico predicted off-target sequences of *Tfap2c* crRNA design 2 by Benchling. red=mismatches

Sequence	PAM	Score	Cut locus	Location		
				Exonic	Intronic	
					Intragenic	Intergenic
AGA <b>A</b> GTGAGGAGAGTG <b>A</b> A GT	AA G	2.4	chr3:+1420882 04			X
AGAC <b>A</b> TGAG <b>C</b> AGAGTGAC <b>T</b> T	AA G	1.1	chr13:- 34148372			X
<b>T</b> G <b>C</b> CG <b>A</b> GTGGAGAGTGAC GT	GG G	0.9	chr4:+1375407 36	<i>Hspg2</i>		
<b>G</b> GACG <b>T</b> AGGAG <b>G</b> GTGAC GT	GA G	0.9	chr18:- 12167392			X
<b>G</b> GAAGTGAG <b>A</b> GGAGTG <b>A</b> CGT	CA G	0.9	chr14:- 106247305			X
<b>G</b> GAAG <b>T</b> GTGGAGAG <b>T</b> GG CGT	CG G	0.8	chr2:+1588439 03			X
AGAG <b>G</b> GTGAG <b>G</b> TGAGTGAC <b>C</b> T	AA G	0.7	chr3:+1025842 31			X
<b>T</b> GAC <b>C</b> T <b>G</b> GGAGAGTGAC TT	CA G	0.7	chr6:- 101005083			X
<b>G</b> GAG <b>C</b> TGAGGAGAGTG <b>A</b> <b>C</b> AT	CA G	0.6	chr3:- 39945780			X
AGAC <b>C</b> TGAG <b>G</b> TGTGTGAC GT	GG G	0.6	chr18:- 46741224			X

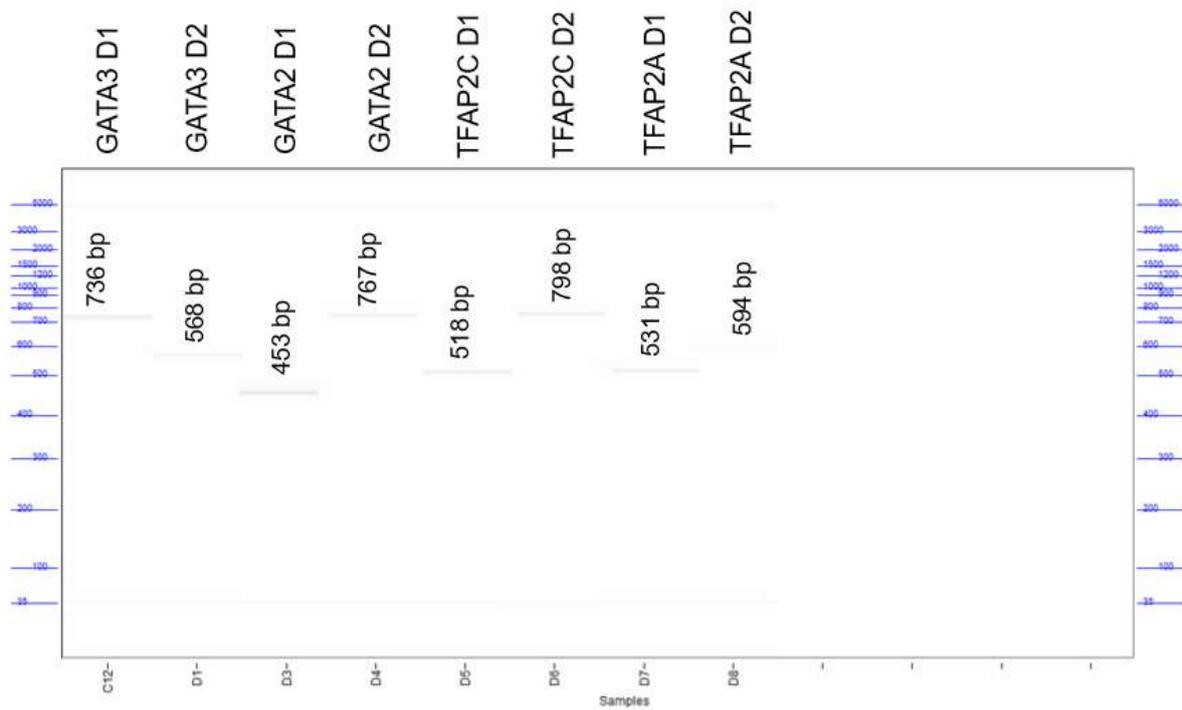
**Supplementary table 13:** Top 10 in-silico predicted off-target sequences of *Tfap2a* crRNA design 1 by Benchling. red=mismatches

Sequence	PAM	Score	Cut locus	Location		
				Exonic	Intronic	
					Intragenic	Intergenic
CGAGGGGACGAGCAACG GGA	CAG	1.7	chr19:+450262 54	<i>Lzts2</i>		
GTCCAGCACCAGCAACG GGA	AGG	1.4	chr2:- 119878106	<i>RP23- 348L16. 1</i>		
CGACTGAAGCAGCAACG GGA	GAG	1.0	chr10:+772950 10			
TTACGGCACGAGCAACCG GA	CAG	0.7	chr16:+207334 97			
CGTGGGTAGCAGCAACG GGA	AGG	0.6	chr5:+5268086 0			
TGCGGCAGCGGCAACG GGA	GAG	0.6	chr4:- 126296960			
CTCGGCAGCAGCAACG GGG	CGG	0.5	chr13:- 36726711	<i>Nrn1</i>		
CGAGGTCCACAGCAACG GGA	AGG	0.5	chr9:- 79966868			
CGGCTGCAGCAGCAACG GGC	TGG	0.5	chr8:- 119698255			
CGCCGGCGCCAGTAACG GGG	AGG	0.3	chr18:+786943 6	<i>Wac</i>		

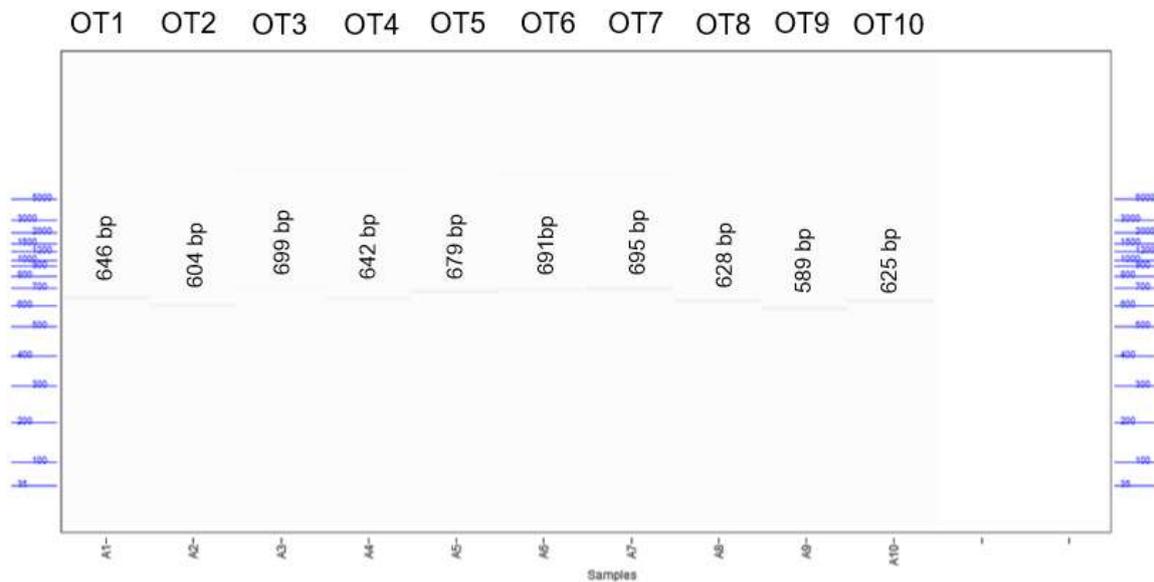
**Supplementary table 14:** Statistical output of Chi square tests performed to compare morphological distributions of KO embryos to Scrambled morphology. S= significant result, NS = non-significant result

Morphological distribution of KO embryos compared to scrambled morphology	p-value ( $\alpha=0.05$ )	Conclusion
<i>Control</i>	0.003	S
<i>Gata3 KO</i>	0.326	NS
<i>Gata2 KO</i>	0.978	NS
<i>Tfap2c KO</i>	0.002	S
<i>Tfap2a KO</i>	0.004	S

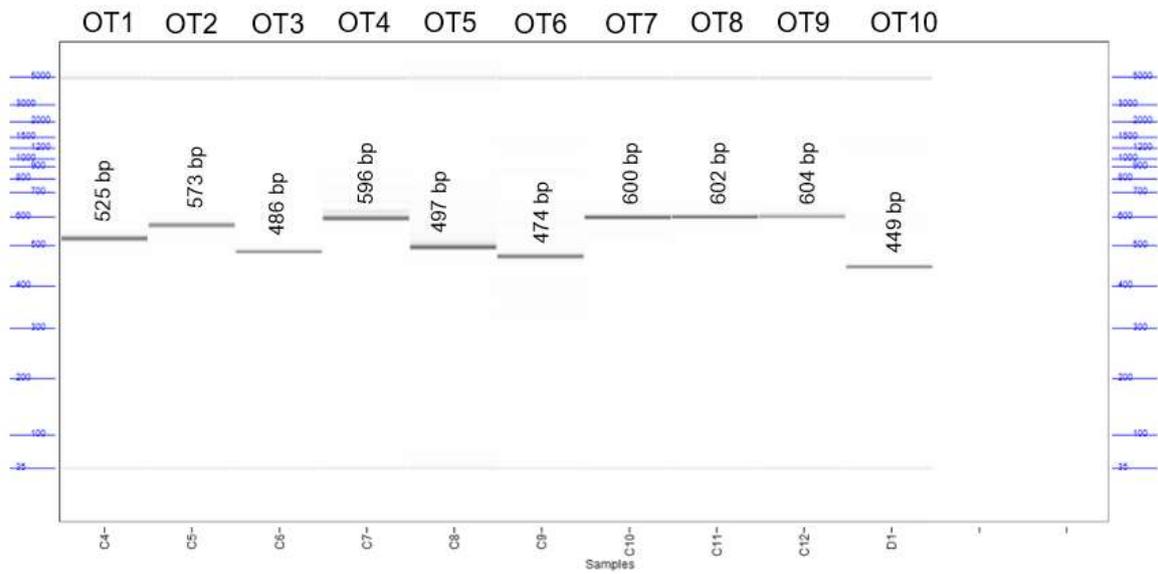
Supplementary figures



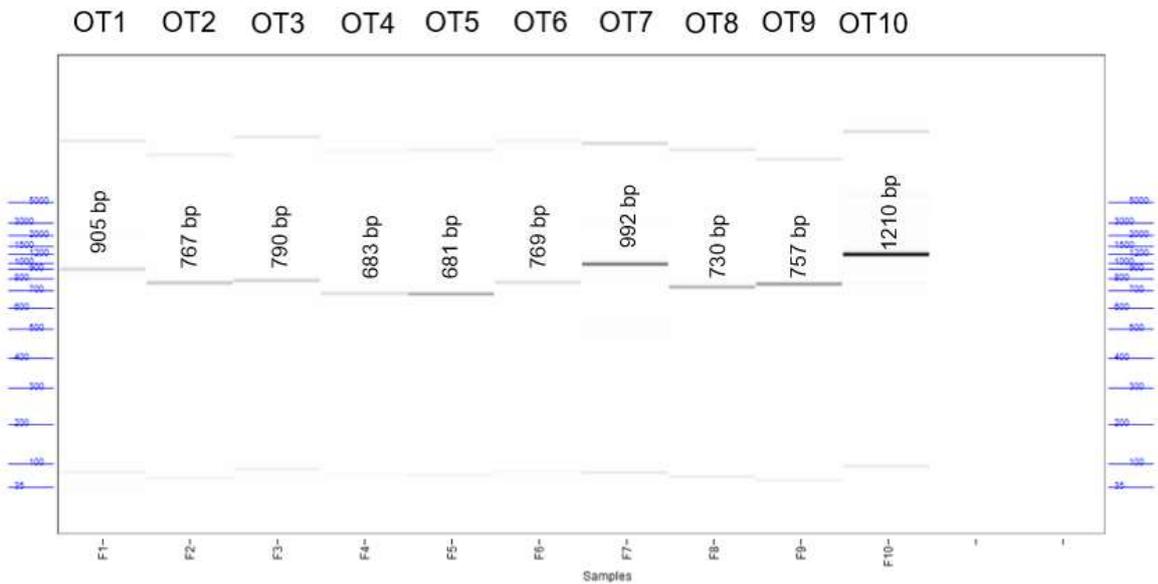
**Supplementary figure 1:** Fragment analyzer gel electrophoresis for each primer pair amplifying the region around the cleavage site per crRNA design for each gene.



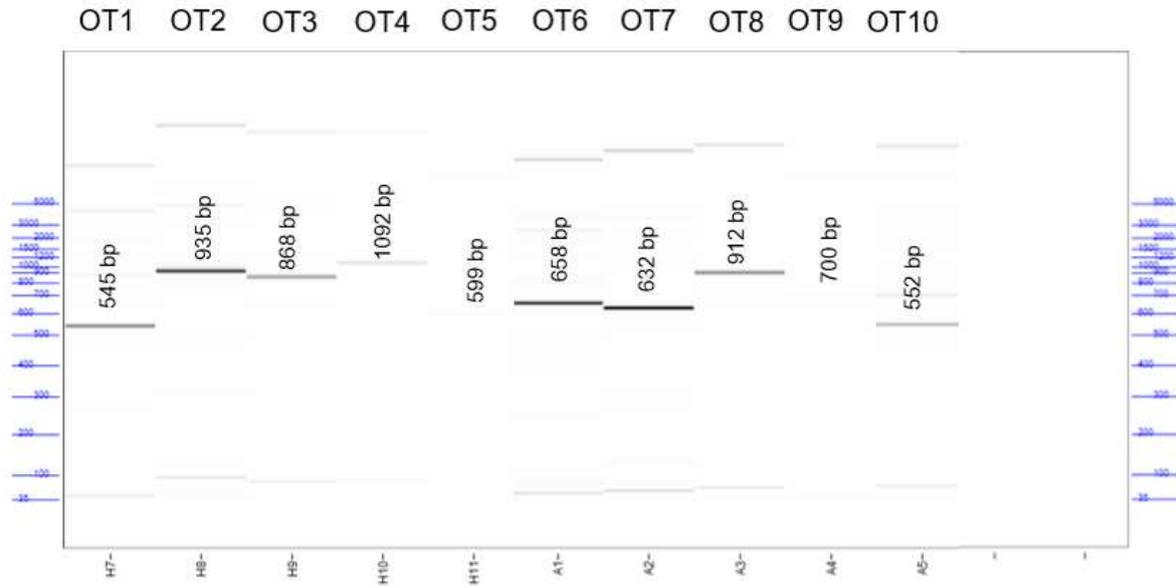
**Supplementary figure 2:** Fragment analyzer gel electrophoresis showing the size (bp) of the amplicons generated of the region surrounding the top 10 in-silico predicted off-target sites of the CRISPR/Cas9 complex containing crRNA design 2 targeting Gata3.



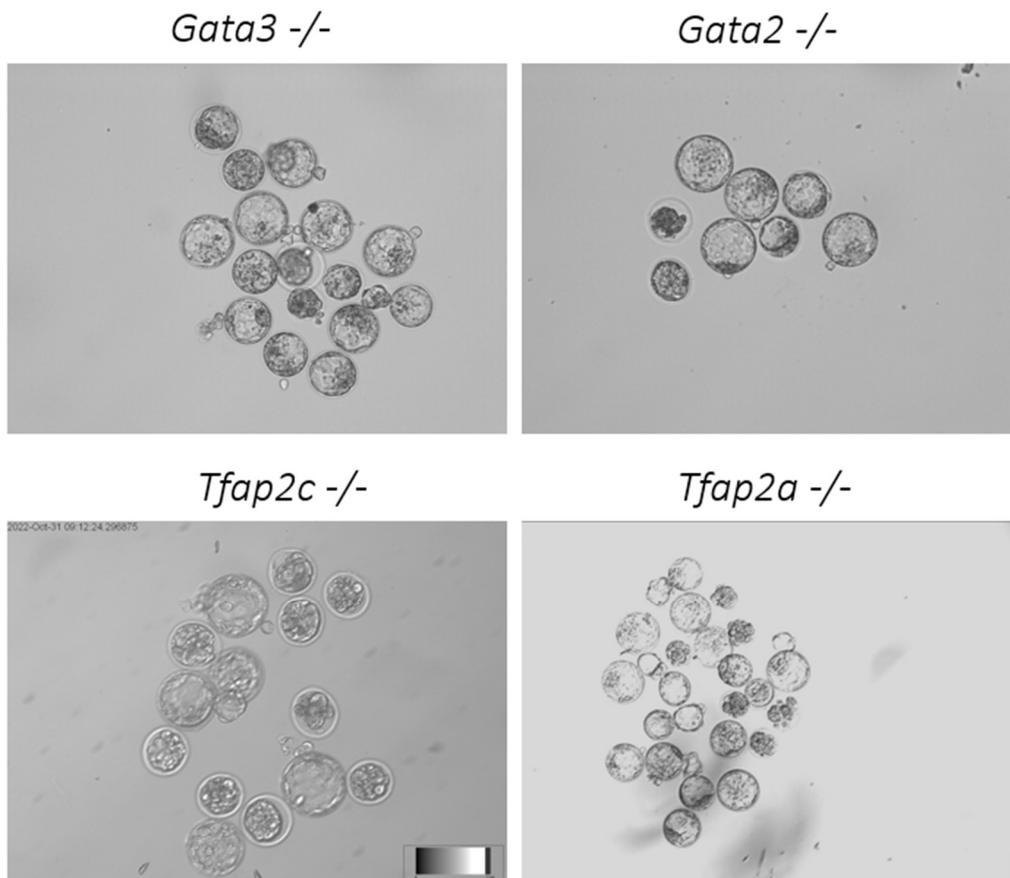
**Supplementary figure 3:** Fragment analyzer gel electrograms showing the size (bp) of the amplicons generated of the region surrounding the top 10 in-silico predicted off-target sites of the CRISPR/Cas9 complex containing crRNA design 2 targeting *Gata2*.



**Supplementary figure 4:** Fragment analyzer gel electrograms showing the size (bp) of the amplicons generated of the region surrounding the top 10 in-silico predicted off-target sites of the CRISPR/Cas9 complex containing crRNA design 2 targeting *Tfap2c*.



**Supplementary figure 5:** Fragment analyzer gel electrophoresis showing the size (number of bps) of the amplicons generated of the region surrounding the top 10 in-silico predicted off-target sites of the CRISPR/Cas9 complex containing crRNA design 2 targeting *Tfap2a*.



**Supplementary figure 6:** Typical embryo morphology of *Gata2/3*  $-/-$  and *Tfap2a/c*  $-/-$  embryos.