

# Understanding Transcriptional Control During Primordial Germ Cell Differentiation

Dissertation

for the award of the degree

“Doctor rerum naturalium” (Dr. rer. nat.)

within the doctoral program *Genes and Development*  
of the Göttingen Graduate Center for Neurosciences, Biophysics,  
and Molecular Biosciences  
at Georg-August-Universität Göttingen,  
Faculty of Biology and Psychology

Submitted by

**Cera Alexandra McDonald**

born in San Diego, California, USA

on April 30<sup>th</sup>, 2022





Thesis committee:

**Dr. Ufuk Günesdogan**, Department of Developmental Biology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology, Georg-August University of Göttingen, Göttingen

**Prof. Dr. Rüdiger Behr**, Platform Degenerative Diseases, German Primate Center, Göttingen

**Dr. Melina Schuh**, Meiosis Department, Max Planck Institute for Multidisciplinary Sciences, Göttingen

Members of the Examination Board:

**1<sup>st</sup> Reviewer: Dr. Ufuk Günesdogan**, Department of Developmental Biology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology, Georg-August University of Göttingen, Göttingen

**2<sup>nd</sup> Reviewer: Prof. Dr. Gregor Bucher**, Department of Developmental Biology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology, Georg-August-University of Göttingen, Göttingen

Further members of the Examination board:

**Dr. Peter Lénárt**, Research Group Cytoskeletal Dynamics in Oocytes, Max Planck Institute for Multidisciplinary Sciences, Göttingen

**Prof. Dr. Ernst Wimmer**, Department of Developmental Biology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology, Georg-August-University of Göttingen, Göttingen

Date of oral exam: June 9<sup>th</sup>, 2022



# Acknowledgements

“El primer puente de Constitución y a mis pies  
Fragor de trenes que tejían laberintos de hierro.  
Humo y silbatos escalaban la noche. . .”  
- Mateo, XXV:30, Jorge Luis Borges

I want to thank first and foremost my supervisor Dr. Ufuk Günesdogan. I have learned so much from you during this time and in the lab, and I am so grateful for all of the wisdom and expertise you've shared with me, as well as all the help and patience in the lead-up to the submission. I am so appreciative of the lab you have created, filled with incredible students and endless possibilities. I am also grateful that the avocado will live on!

I sincerely want to thank to my TAC committee members Prof. Rüdiger Behr and Dr. Melina Schuh for their insight, wisdom, and kindness over these last four years.

Many, many thanks to thank Prof. Ernst Wimmer for always providing help, guidance, and incredible stories, as well as Prof. Gregor Bucher for wonderful advice.

I would further like to thank the members of the developmental biology department, but especially Dr. Nico Posnien, Merle Eggers, and Angelika Löffers. Thanks as well to Bettina Hucke, Beate Preitz, Claudia Hinnert, Elke Küster, Helma Gries, Vera Terblanche, and Katrin Kanbach, who the department could not run without and have been exceptional colleagues (and very patient with my terrible German).

Profound thanks to Xiaojuan Li and Ignacio Rodríguez-Polo. I do not know what I would have done without your help, and I have no words to express how much your kindness, time, and patience meant to me, I am so lucky to have worked with you both.

I want to express my thanks and deep gratitude to my labmates Dominik Mühlen, Erica Calabrese, Shyam Sundar Alagarsamy Ramasamy, and Oleksandr Dovgusha. You have all been incredible colleagues, friends, and scientists, and you make wherever you are a better place to be.

I would like to thank all of the students, post-docs, and former members of the department, but special thanks to Hassan Mutasim Mohammed Ahmed, Musa Danazumi Isah, and Bibi Atika, who were so kind to me when I joined, and helped me when I was most lost. You have

been incredible friends and I am so lucky to have met you. Additional thanks to Constanza Tapia Contreras, Ting-Hsuan Lu, Muhammad Salim Hakeemi, and Hazem Khalifa for being wonderful colleagues and friends.

Endless love and gratitude to my parents, Julie McDonald and Gerry McDonald, as well as my grandmother Patricia McDonald. I am so profoundly lucky to have you as my family, and words cannot express my thankfulness for your unceasing love and support. Love and thanks to Daisy, Liam Neeson, Izzy, Sitka, Annie, Harriet, and Harold.

I wanted to thank Chris Ehrlich and Jocelyn Pearce, as I could not ask for truer friends. I can't wait to see you both again on the other side.

Sincerest thanks to Sarah Nakano Purgett, Sam Anklesaria, Rachel Levine, Alex Hagen, Nina Hagen, Charlie Bendelsmith, Greta Stacy, Madelyn Olson, Lydia Somers, and Dylan Saul. I am so lucky to have known you all, and to be able to be your friend.

Julian Schwanbeck - words fail me, but I think you understand. What kismet to have met you at all. To large oranges, boulevardiers, rooftop coffees, and being "the only warm thing for miles & the only thing that can't shine". Jeremy Bearimy, baby.

Although I am not sure they'll ever see this, I nevertheless wanted to thank the people who shaped how I see the world, and who themselves have brought the world only good: Gunn Mangerud, Kari Kunze, Jim Schwieger, Maureen Conway, Renae Wantock, Jose Reyna, Tho Bui, Avalon Levey, Diana Popescu, Annette Eneanya, Archishman Sarkar, Clara Lücke, Paola Solimena, Lorin Diaconu, Vlad Kim, Dianna Bautista, Adi Mackay, Sevgi Cengiz, Gökçen Gözüim, Ahmet Bakırbaş, Nagwa Mohamedy, Rania Mohamed Younis, Rasha Mohamed Younis, Rubi Rabach, Brian McDonald, LaNor McDonald, Jan Swanson, Sunna Rothmaler, Sarah Coutinho Saraiva, Pasqualina Legorano, Corinna Kulicke, and Clara Miliotis.



# Contents

<b>Acknowledgements</b> . . . . .	<b>1</b>
<b>Summary</b> . . . . .	<b>9</b>
<b>Introduction</b> . . . . .	<b>11</b>
1.1 Embryonic development in the mouse . . . . .	11
1.1.1 Pre-implantation development . . . . .	11
1.1.2 Post-implantation embryonic development . . . . .	13
1.2 Primordial germ cells (PGCs) . . . . .	15
1.2.1 PGC specification and maturation . . . . .	16
1.2.2 Key PGC factors . . . . .	18
1.2.2.1 “Tripartite transcription factor network” . . . . .	18
1.2.2.2 AP2 $\gamma$ ( <i>Tfap2c</i> ) . . . . .	18
1.2.2.3 BLIMP1 ( <i>Prdm1</i> ) . . . . .	19
1.2.2.4 PRDM14 ( <i>Prdm14</i> ) . . . . .	19
1.2.2.5 Overlapping roles of PGC and naïve pluripotency markers . . . . .	20
1.2.3 <i>In vitro</i> recapitulation of PGC induction . . . . .	20
1.3 Transcription factor AP2 $\gamma$ . . . . .	21
1.3.1 AP2 $\gamma$ overview . . . . .	22
1.3.2 AP2 $\gamma$ in early embryonic development . . . . .	25
1.3.3 AP2 $\gamma$ in mammary development . . . . .	26
1.3.3.1 Mammary gland development in the mouse . . . . .	26
1.3.3.2 AP2 $\gamma$ in mammary gland development . . . . .	27
1.4 <i>Cis</i> -regulatory elements . . . . .	27
1.4.1 Identifying and classifying CREs . . . . .	28
1.4.2 Functional testing of enhancers with CRISPR/Cas9 . . . . .	31
1.4.3 CRISPR/Cas9 genetic screens . . . . .	32
1.5 Scope and aims of work . . . . .	33
<b>Materials and methods</b> . . . . .	<b>35</b>
2.1 Cell culture . . . . .	35
2.1.1 ESC culture . . . . .	35
2.1.2 EpiLC and PGCLC induction . . . . .	35
2.1.3 Transfection of ESCs . . . . .	37
2.1.4 293T-HEK Culture . . . . .	38

2.2	Generation of genetically-modified ESC lines . . . . .	38
2.2.1	Single-gRNA CRISPR/Cas9-based KO lines . . . . .	38
2.2.2	Homology-directed repair plasmids . . . . .	41
2.3	FACS . . . . .	44
2.4	qRT-PCR . . . . .	46
2.5	Indirect immunofluorescence . . . . .	48
2.5.1	IF staining of cell monolayers . . . . .	48
2.5.2	IF staining of cryosections . . . . .	48
2.5.3	Image analysis . . . . .	49
2.6	Western blotting . . . . .	49
2.7	RNA-seq . . . . .	51
2.7.1	RNA-seq library preparation and sequencing . . . . .	51
2.7.2	RNA-seq data analysis . . . . .	52
2.8	ATAC-seq . . . . .	53
2.9	<i>Tfap2c</i> tiling screen . . . . .	56
2.9.1	<i>Tfap2c</i> -screen gRNA library design . . . . .	56
2.9.2	<i>Tfap2c</i> -screen gRNA library generation . . . . .	57
2.9.3	Generation of lentivirus . . . . .	61
2.9.4	<i>Tfap2c</i> tiling screen in ESCs, EpiLCs, and PGCLCs . . . . .	63
2.9.5	Library preparation, sequencing, and analysis of <i>Tfap2c</i> screen results . . . . .	64
	<b>Results . . . . .</b>	<b>67</b>
3.1	Knockout of AP2 $\gamma$ in mouse ESCs . . . . .	67
3.2	RNA-sequencing of AP2 $\gamma$ -KO cells . . . . .	71
3.2.1	Determination of differential expression in ESCs, EpiLCs, and d4/d6 GFP+ cells . . . . .	71
3.2.2	Upregulation of EMP markers in AP2 $\alpha$ -KO d4/d6 GFP+ cells . . . . .	83
3.2.3	Validation of the AP2 $\gamma$ -KO phenotype in an SGET ESC line . . . . .	85
3.3	Analysis of chromatin accessibility in AP2 $\gamma$ -KO cells . . . . .	90
3.3.1	TSS accessibility examined in AP2 $\gamma$ -KO cells . . . . .	90
3.3.2	Characterization of DAPs upon AP2 $\gamma$ -KO . . . . .	91
3.3.3	Expression of markers of neural fate in AP2 $\gamma$ -KO cells . . . . .	103
3.4	<i>Tfap2c</i> Tiling screen . . . . .	105
3.4.1	Development of AP2 $\gamma$ reporter cell line . . . . .	105
3.4.2	<i>Tfap2c</i> tiling screen design and preparation . . . . .	110
3.4.3	Tiling screen of <i>Tfap2c</i> TAD in SGET-Cas9 cells . . . . .	112
	<b>Discussion . . . . .</b>	<b>119</b>

4.1	The role of AP2 $\gamma$ in PGC induction . . . . .	119
4.1.1	Loss of AP2 $\gamma$ affects the morphology of ESCs, but not that of EpiLCs or EBs . . . . .	120
4.1.2	AP2 $\gamma$ -KO results in misregulation of imprinted genes and dramatically alters chromatin accessibility . . . . .	121
4.1.3	AP2 $\gamma$ -KO EpiLCs upregulate lineage non-specific markers . . . . .	124
4.1.4	Loss of AP2 $\gamma$ results in impaired PGCLC differentiation . . . . .	125
4.1.5	Chromatin accessibility is altered during PGCLC induction from AP2 $\gamma$ -KO EpiLCs . . . . .	127
4.1.6	Different effects of AP2 $\gamma$ -KO on ESCs and d4 GFP <sup>+</sup> cells . . . . .	128
4.1.7	Conclusions on the study of AP2 $\gamma$ activity . . . . .	130
4.1.7.1	AP2 $\gamma$ should be evaluated as a potential factor supporting naïve pluripotency . . . . .	130
4.1.7.2	The characterization of AP2 $\gamma$ as a potential pioneer factor is not complete . . . . .	133
4.2	Single gRNA tiling screen of <i>Tfap2c</i> TAD identifies single gRNAs of interest . . . . .	133
4.2.1	Basis of screen . . . . .	133
4.2.2	gRNA enrichment is demonstrated in each examined cell type . . . . .	134
4.2.2.1	gRNAs primarily target non-coding regions in ESCs . . . . .	135
4.2.2.2	Strong negative enrichment of gRNAs targeting exons in EpiLCs . . . . .	137
4.2.2.3	gRNA enrichment in PGCLCs is primarily positive . . . . .	138
4.2.3	Conclusions on <i>Tfap2c</i> tiling screen . . . . .	139
4.2.3.1	Identification of single gRNAs for further characterization . . . . .	139
4.2.3.2	Improvements suggested for further screens . . . . .	140
	<b>Bibliography . . . . .</b>	<b>143</b>
	<b>Appendix . . . . .</b>	<b>i</b>
7.1	Additional buffers used in this work . . . . .	i
7.1.1	Gibson assembly mastermix . . . . .	i
7.1.2	LB Agar . . . . .	ii
7.1.3	PBS, 10x stock . . . . .	ii
7.1.4	TBE, 10x stock . . . . .	ii
7.2	Antibodies . . . . .	iii
7.3	Cell culture media . . . . .	iv
7.3.1	2i+LIF medium . . . . .	iv
7.4	Cell lines used in this study . . . . .	vi
7.5	Computer programs and bioinformatics tools . . . . .	vii



7.6	Gels . . . . .	viii
7.6.1	SDS-PAGE gels . . . . .	viii
7.6.2	TBE gel, 10% . . . . .	viii
7.7	Materials . . . . .	ix
7.7.1	Cell culture . . . . .	ix
7.7.2	Chemicals and reagents . . . . .	xiii
7.7.3	Disposables . . . . .	xv
7.7.4	Enzymes and mastermixes . . . . .	xvi
7.7.5	Equipment . . . . .	xvii
7.7.6	Kits . . . . .	xviii
7.8	Plasmids . . . . .	xix
7.8.1	2C::tdTomato Reporter . . . . .	xix
7.8.2	phNANOS3-T2A-tdTomato-RoxPGKPuro . . . . .	xx
7.8.3	pCAGGS-Dre-IH . . . . .	xxi
7.8.4	pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W . . . . .	xxii
7.8.5	pMD2.G . . . . .	xxiii
7.8.6	pmTfap2c-T2A-tdTomato-RoxPGKPuro . . . . .	xxiv
7.8.7	pPB-CAG-Cas9-IRES-Hygro . . . . .	xxiv
7.8.8	pPY-CAG-mKO-Puro-IP . . . . .	xxiv
7.8.9	pPY-CAG-PBase . . . . .	xxiv
7.8.10	psPAX2 . . . . .	xxv
7.8.11	pX330-U6-Chimeric_BB-CBh-hSpCas9 . . . . .	xxvi
7.9	Sequencing quality control and supplementary data . . . . .	xxvii
7.9.1	ATAC-sequencing . . . . .	xxvii
7.9.1.1	ATAC-seq qualitative library assessment . . . . .	xxvii
7.9.1.2	ATAC-seq library fragment length distribution . . . . .	xxviii
7.9.1.3	ATAC-seq library transcription start sites . . . . .	xxix
7.9.1.4	ATAC-seq PCA . . . . .	xxx
7.9.1.5	Differentially accessible peaks between WT cell types . . . . .	xxx
7.9.1.6	Example DAPs of interest identified in ATAC-seq libraries . . . . .	xxx
7.9.1.7	Genes annotated to DAPs . . . . .	xxxii
7.9.2	<i>Tfap2c</i> tiling screen sequencing . . . . .	xxxiii
7.9.2.1	Quality control statistics of <i>Tfap2c</i> tiling screen . . . . .	xxxiii
7.9.2.2	Expression of genes covered in <i>Tfap2c</i> tiling screen . . . . .	xxxiv
7.9.3	RNA-sequencing . . . . .	xxxv
7.9.3.1	Direct comparisons of RNA-sequencing libraries . . . . .	xxxv
7.9.3.2	Expression of SOX TF family genes with development following KO of AP2 $\gamma$ . . . . .	xxxvii

7.9.3.3	Expression of TEAD TF family genes with development following KO of AP2 $\gamma$ . . . . .	xxxviii
7.9.3.4	IGV visualization of cell line KO3 . . . . .	xxxix
7.10	Synthetic DNA sequences . . . . .	xl
7.10.1	Oligos . . . . .	xl
7.10.1.1	CRISPR/Cas9 knockout oligos . . . . .	xl
7.10.1.2	pmTfap2c Gibson assembly oligos . . . . .	xli
7.10.2	Primers . . . . .	xlii
7.10.2.1	Genotyping and plasmid-sequencing primers . . . . .	xlii
7.10.2.2	NGS sequencing primers . . . . .	xlvi
7.10.2.3	qPCR primers . . . . .	xlix
7.10.3	gBlocks . . . . .	l
7.11	<i>Tfap2c</i> tiling screen gRNA design . . . . .	li
<b>List of Figures . . . . .</b>		<b>liv</b>
<b>List of Tables . . . . .</b>		<b>lxiv</b>
<b>Glossary . . . . .</b>		<b>lxvii</b>

# Summary

AP2 $\gamma$ , encoded by *Tfap2c*, is a transcription factor which both represses and activates target gene expression. It is required for the survival of the early mouse embryo; depending on the context of its depletion, the loss of the protein may prevent the correct formation of blastocysts, result in placental defects and embryonic death, or result in infertility of the adult mouse. AP2 $\gamma$  expression has been identified in the maintenance of adult tissues, and its expression has been localized to the skin, brain, and mammary gland. It has also been identified in numerous cancers, as a prognostic marker of both poor and good outcomes, depending upon the cancer and context. It is thus a fascinating transcription factor, as its expression or loss thereof can tip the scales of survival in both the embryo and adult.

AP2 $\gamma$  was examined in this work in the context of its role as a transcription factor required for primordial germ cell (PGC) differentiation. PGCs are unipotent embryonic cells which give rise to either sperm or egg, and at the point of their induction are the only embryonic cells which will be able to transmit heritable genetic and epigenetic information. The knockout (KO) of AP2 $\gamma$  has been shown to impair PGC differentiation and result in sterility of otherwise healthy, adult mice. However, the transcriptional regulation exerted by AP2 $\gamma$  during PGC differentiation is poorly understood. One hypothesis for the strong transcriptional effect exerted by AP2 $\gamma$  is that it may be a pioneer factor, a protein which is capable of binding to and opening closed chromatin to allow other transcription factors and adapter proteins to bind and thus effect transcriptional regulation, however it is still unknown if AP2 $\gamma$  is truly a pioneer factor, and if so in which of its many biological contexts does it exert this effect.

My work aimed to improve our understanding of the role of AP2 $\gamma$  in PGC differentiation in two ways: First, to characterize the role of AP2 $\gamma$  as a regulator of transcriptional activity and as a potential pioneer factor in embryonic stem cells (ESCs), epiblast-like cells (EpiLCs), and PGC-like cells (PGCLCs). These cells represent the *in vitro* recapitulation of the developmental process which allows for the differentiation of ESCs towards PGCLC fate, and thus allows for the examination of transcriptional events resulting in the correct differentiation of PGCLCs. To accom-

plish this, I generated AP2 $\gamma$  knockout (KO) ESCs and I applied RNA-sequencing (ribonucleic acid sequencing) transcriptomic analysis to identify changes in gene expression as I differentiated the cells towards PGCLCs. I was able to show that, in the absence of AP2 $\gamma$ , PGCLCs cannot be correctly specified and further, numerous marker genes of naïve pluripotency and somatic fate are misregulated, raising broader questions about the role of AP2 $\gamma$  not only in PGCLCs but also within the context of naïve pluripotent ESCs. I also examined the chromatin accessibility of the AP2 $\gamma$ -KO cells as they progressed from ESCs towards PGCLC fate by performing ATAC-sequencing (assay for transposase-accessible chromatin with high-throughput sequencing). I was able to identify differentially accessible chromatin in all examined cell types, however most dramatically in the AP2 $\gamma$ -KO ESCs, further underlining the potential role of AP2 $\gamma$  in naïve pluripotency.

My second aim was to identify *cis*-regulatory elements (CREs) including enhancers, regulating the expression of AP2 $\gamma$  as the cells differentiated from ESCs towards PGCLC fate. To this end I employed a lentiviral CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/Cas9) based screen approach, tiling the topologically-active domain of AP2 $\gamma$  with gRNAs to KO potential regulatory elements acting on the gene. Analysis of the screen results did not clearly indicate CREs regulating AP2 $\gamma$  expression, nevertheless interesting genomic sites were identified which may be investigated further as potential cell type-specific regulatory regions.

This work broadens our understanding of AP2 $\gamma$  in the context of naïve pluripotent ESCs as well as during the process of PGCLC differentiation. It also raises questions about the role of AP2 $\gamma$  as a pioneer factor, and the relationship between AP2 $\gamma$  and the closely related protein AP2 $\alpha$ . In addition, although the employed screen was not able to identify specific CREs driving AP2 $\gamma$  expression, I was able to identify interesting targets of further investigation including potential regulatory regions involved in cell survival during the process of differentiation towards PGCLC fate.

# Introduction

## 1.1 Embryonic development in the mouse

The formation of the totipotent zygote, one single cell resulting from the fertilization of egg by sperm and capable of giving rise to all embryonic and extra-embryonic lineages [1], represents the first moment of creation for a new life. From this point on, a highly regulated and meticulously timed developmental program will coordinate the growth, differentiation, and maturation of the multitudinous tissues and specialized cells required to produce and sustain adult life. Misregulation of development in the maturing embryo can prove catastrophic to the induction or differentiation of tissues, and can even result in miscarriage of pregnancy [2]. Thus, understanding early development not only improves our knowledge of the beginnings of life, but also allows us to investigate and potentially ameliorate errors in development, which in the context of this work concern miscarriage or infertility.

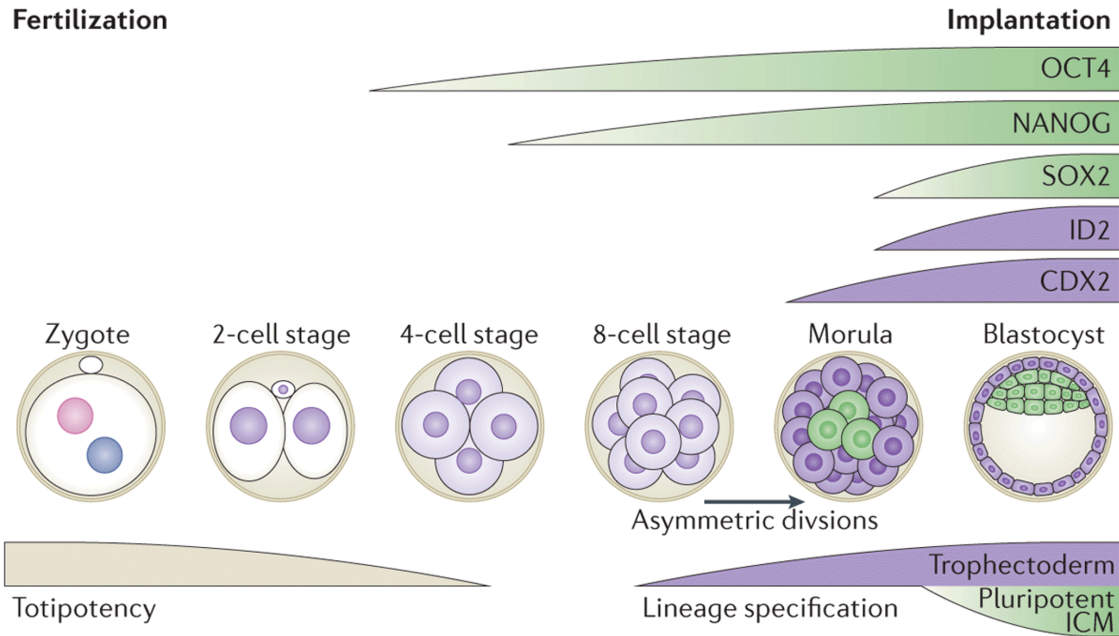
### 1.1.1 Pre-implantation development

Early embryonic development in the mouse (*Mus musculus*) begins with the fertilized zygote, which contains two pronuclei, each corresponding to the segregated paternal and maternal genomes [1] (Figure 1). After fertilization, the pronuclei unite and the embryo divides and enters the two-cell stage. At this point zygotic gene transcription begins to overtake the transcription of remnant maternal messenger ribonucleic acids (mRNAs) in a process termed zygotic genome activation (ZGA) [3].

At the four-cell stage the embryonic cells are not morphologically distinct from each other, nevertheless variations have been identified in transcriptional and epigenetic features between the cells, *e.g.* the differential expression of the pluripotency factor genes *Prdm14* (PR/SET domain 14) [5] and *Nanog*<sup>1</sup> [7]. The significance of this heterogeneous expression is not known, as until the eight-cell stage all cells in the early embryo can give rise to both embryonic and extraembryonic lineages [1]. By the eight-cell stage the blastomere is polarized, meaning inherent polar properties, *e.g.* compositional differences in the apical versus basolateral membranes, may now be ascribed to the cells [1]. Polarization initiates uniform divergence in cell fate, as the hippo signaling pathway is active in the apolar cells of the inner embryo and inactive in the polarized outer cells, resulting in differential gene expression between the two cell populations [8]. Between the eight- and sixteen-cell stages embryonic

---

<sup>1</sup>Nanog derives from the old Irish “Tír na nÓg”, meaning “Land of Youth” *i.e.* the Celtic Otherworld [6].



**Figure 1: Early embryonic development in the mouse.** Totipotency is established in the zygote following fertilization, but is lost once asymmetric cell division produces the distinct inner and outer cell populations of the morula by E2.5. Induction of the trophectoderm (TE; indicated in purple) begins in the morula stage with the formation of polarized, compacted cells which begin to express TE factor CDX2. Induction of the pluripotent inner cell mass (ICM) also begins in the morula with the formation of apolar cells which begin to express pluripotency factors OCT4 and NANOG. By E3.5 in the blastocyst, the TE has formed and begins to upregulate TE factor ID2. At this stage the ICM has also formed, and begins to express pluripotency factor SOX2. Modified from Burton & Torres-Padilla, 2014 [4].

compaction occurs, and is the process by which the cells' morphology is altered; flattening and increasing cell-cell contact [9].

The sixteen-cell morula arises by embryonic day (E) 2.5 from asymmetric cell division, which results in two discrete populations of inner and outer cells (Figure 1) [4]. The development of the blastocyst by E3.5 represents the first symmetry breaking event of the early embryo, as the inner cell mass (ICM) and trophectoderm (TE) are completely specified [8]. The ICM derives from the inner cells of the morula, and by E4.5 will have further differentiated into either the pre-implantation epiblast or the primitive endoderm (PrE) [10]. The epiblast will ultimately develop into the embryo proper and characteristically expresses marker genes *Pou5f1* (POU class 5 homeobox 1, encoded by OCT4), *Nanog*, and *Sox2* (SRY (sex determining region Y)-box 2) [10]. The PrE participates in the formation of the yolk sac and also pro-

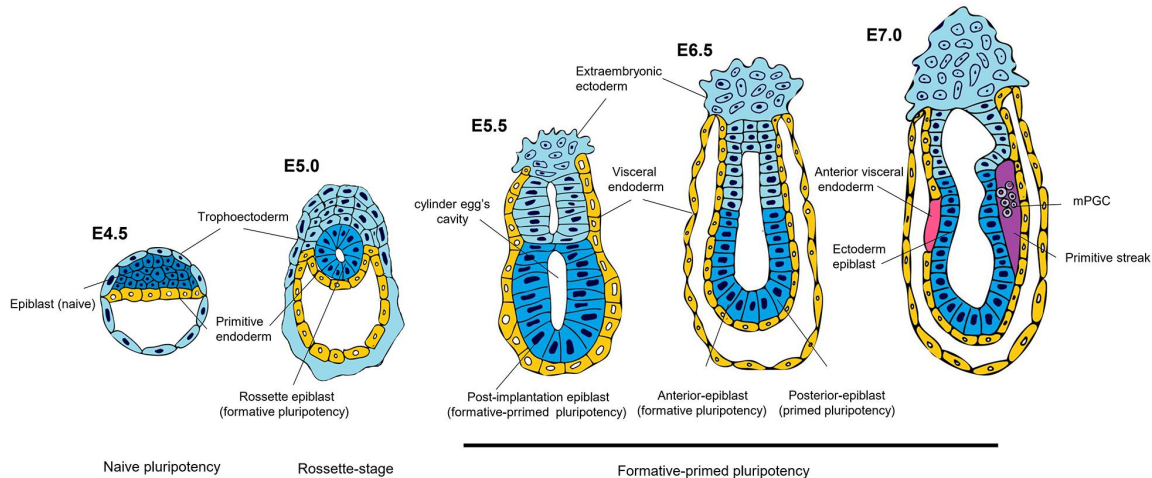
vides spatial separation between the epiblast and the blastocyst cavity [8], and is indicated by the expression of marker genes *Gata6* (GATA binding protein 6 ) and *Pdgfra* (Platelet-derived growth factor receptor A ) [10]. The TE will ultimately comprise the placenta [1], and is indicated by the expression of marker genes including *Cdx2* (Caudal type homeobox 2), *Tead4* (TEA domain transcription factor 4), *Tfap2c* (Transcription factor AP2 gamma), *Eomes* (Eomesodermin), and *Gata3* (GATA binding protein 3) [10]. At this point, cells which have achieved epiblast, PrE, or TE fate cannot exit their lineage and join another lineage *in vivo* [10].

Naïve pluripotency - a developmentally naïve “ground state” [11], an epigenetic “blank slate” [11], simply “sublime” [12] - describes the developmental potential of epiblast cells of the pre-implantation embryo, which are competent to give rise to all embryonic lineages [12]. The naïve state is in part achieved by the global demethylation of the parental genomes which is initiated immediately after fertilization [13]. By E4.5, the cells of the ICM are globally hypomethylated [13], thus allowing access to and transcription of genomic regions which are epigenetically restrained in differentiated cells. To study this cellular state *in vitro*, cells harvested from the murine E4.5 pre-implantation epiblast are cultured in 2i+LIF medium; a defined medium containing GSK3 (Glycogen synthase kinase 3) inhibitor CHIR99021 and MEK (Mitogen-activated protein kinase) inhibitor PD0325901 (two inhibitors, “2i”), as well as LIF (Leukemia inhibitory factor) [14]. These cells are termed embryonic stem cells (ESCs), and are able to contribute to a chimera and the germline when injected into a pre-implantation blastocyst [11].

### 1.1.2 Post-implantation embryonic development

The implantation of the embryo into the uterine wall from E4.75-5.0 [15] represents an active process between an implantation-competent blastocyst and a receptive uterus [16]. Once the blastocyst and the uterus are in physical contact, trophoblasts from the nascent TE invade and colonize the uterine wall [16]. In parallel, the uterine stroma responds to estrogen and progesterone signaling to produce decidual cells in a process known as decidualization [16]. The implanted E5.0 epiblast rosette (Figure 2) is considered to have “formative” pluripotency [12].

By E5.5, the post-implantation embryo initiates a program of differentiation which affects all embryonic and extraembryonic tissues. This process includes cavitation, which forms the proamniotic cavity and reshapes the embryo into an egg cylinder



**Figure 2: Post-implantation development in the early mouse embryo.** The developmental timeline of the mouse embryo is shown from E4.5-7.0. Cells transition from naïve pluripotency at E4.5 through formative pluripotency by E5.0 and primed pluripotency of the posterior-epiblast by E6.5. Cells deriving from the epiblast lineage and which will ultimately give rise to the embryo proper are shown in dark blue. Cells of the TE lineage which will ultimately form or contribute to the placenta are shown in light blue. Cells deriving from the PrE lineage and which give rise to the visceral endoderm (VE) are shown in yellow. The anterior visceral endoderm (AVE) is indicated in pink, while the primitive streak is indicated in purple. The emergence of the primordial germ cells (PGCs) by E7.0 is indicated. Figure taken from Ávila-González *et al.*, 2021 [12].

[17] (Figure 2). The polar embryonic TE differentiates into the extra-embryonic ectoderm (ExE) and harbors the trophoblast stem cells (TSCs) which will later form the placenta, while the mural abembryonic TE produces trophoblast giant cells (TGCs) [15], terminally differentiated cells which assist in implantation and placentation [18]. The primitive endoderm differentiates into the parietal endoderm (PE), which associates with the TGCs, and the visceral endoderm, which envelops the developing embryo [15]. At E5.5 the visceral endoderm is comprised of three types of tissue: the extraembryonic visceral endoderm (ExVE), the embryonic visceral endoderm (EmVE), and the distal visceral endoderm (DVE) [17].

The E5.5 epiblast cells exhibit formative pluripotency, a transitional stage in which the cells are becoming primed for lineage differentiation [19]. Formative pluripotent cells can contribute to a chimera if they are injected into a blastocyst, however their contribution is “patchy” and not well distributed [20]. This is in part due to the fact that *de novo* methylation resumes in the epiblast following implantation, and continues until E6.0 when the embryonic methylation pattern is fully established [13]. Thus the formative pluripotent cells face epigenetic limitations on their ability to access and express the full library of genes which are required to maintain naïve

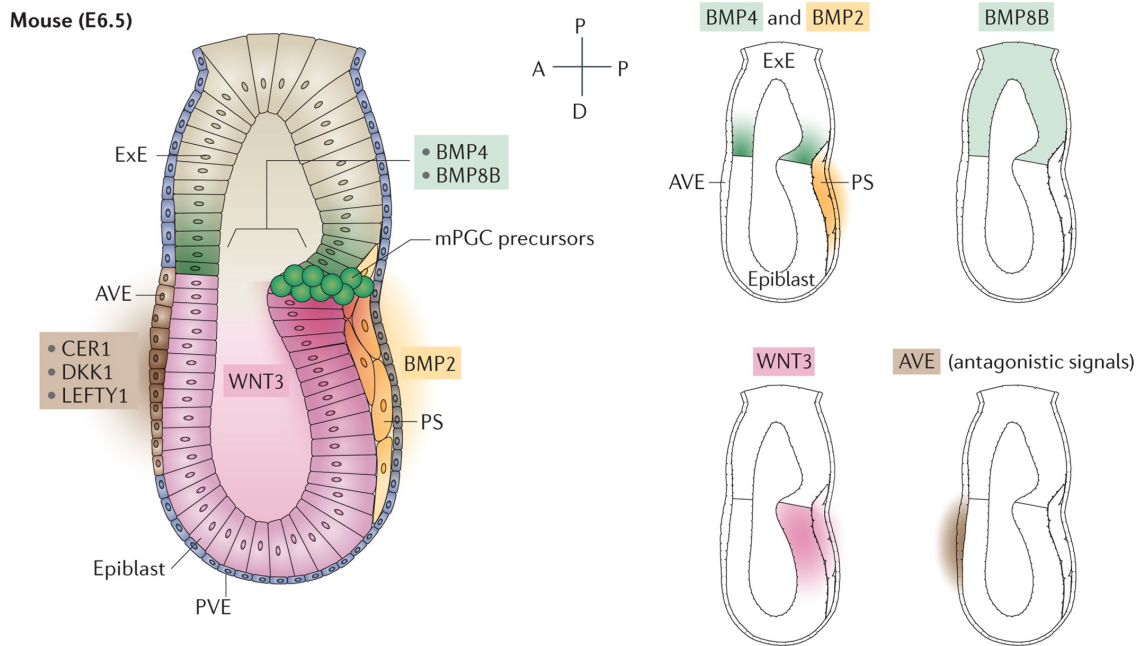


pluripotency. A self-renewing population of epiblast cells known as epiblast-derived stem cells (EpiSCs) can be harvested from the E5.5-8.0 embryo and raised in medium containing bFGF (basic fibroblast growth factor) and Activin A [11]. These cells are considered to exhibit primed pluripotency [19], as they are fully primed for lineage differentiation. Additionally, a transient population of epiblast-like cells (EpiLCs) akin to cells of the E5.5 epiblast may be induced *in vitro* directly from ESCs, a process which also requires the addition of bFGF and Activin A in medium containing 1% knockout serum replacement (KSR) [11; 21].

Gastrulation begins by E6.5, and is the process by which the embryo forms multicellular germ layers, develops central axes, and rearranges the location of cellular populations to establish a multilayered, complex gastrula [22]. The primitive streak also appears at this time point, and is a short-lived structure which helps to establish bilateral symmetry and the central axes [23]. During this dynamic period, transcriptional differences have been identified between the anterior and posterior epiblast [12]. It may be that regionalized pluripotency is in effect during this time, such that the anterior epiblast lags behind the posterior epiblast by continuing to exhibit formative pluripotency while the posterior epiblast has already entered the state of primed pluripotency (Figure 2) [12]. These differences in localized pluripotency may help to provide further spatial and structural segregation in the growing embryo.

## 1.2 Primordial germ cells (PGCs)

By the onset of gastrulation, a small population of primordial germ cell (PGC) precursors emerges in the E6.5 embryo [24]. The induction and differentiation of PGCs occurs in parallel with the development of the embryo, and at the time of their induction, PGCs are the only embryonic cells which will transmit heritable genetic and epigenetic information to offspring [19]. PGCs have thus been described as an “eternal link between generations” [25]. As unipotent cells, PGCs exist to ultimately differentiate into mature germ cells and confer fertility to the adult mouse. Unraveling the complex processes and identifying the key factors necessary for correct PGC fate allows one to better understand when and how misregulation affects fertility.



**Figure 3: PGC induction in the mouse.** Mouse PGC precursors (mPGCs; shown in green) are induced following exposure to activating signals BMP4 and 8B (shown in green) from the extraembryonic ectoderm (ExE; shown in beige), BMP2 signals from the proximal visceral endoderm (VE; shown in yellow), and WNT3 signals from the posterior epiblast and posterior VE (both shown in pink). Inhibitory signals including CER1, DKK1, and LEFTY1 released from the anterior VE (AVE; shown in brown) prevent the induction of PGCs, and thus control the location and opportunity for PGC induction. Together, these activating and antagonizing signals ensure a small population of mPGCs may be specified in the proximal-posterior region of the epiblast. The directional axes of the embryo are indicated with P - D (proximal to distal) and A - P (anterior to posterior). The primitive streak (PS) is also indicated. Modified from Tang *et al.*, 2016 [24].

### 1.2.1 PGC specification and maturation

In the mouse, precursor PGCs are specified from somatic cells on E6.25 in the proximal-posterior epiblast [26], and their induction requires a coordination of signals from the surrounding tissues (Figure 3) [24]. In order for a somatic cell to be induced towards PGC fate, it must receive BMP4 [26; 27; 28; 29] and BMP8B [30] signals from the ExE as well as BMP2 signals from the proximal VE [26; 29]. BMP signaling results in the phosphorylation of SMAD (Suppressor of mothers against decapentaplegic) proteins 1, 5, and 8, and the phosphorylated proteins dimerize with SMAD4 to enter the nucleus and begin to regulate target genes [31; 32]. WNT3 (Proto-oncogene protein Wnt-3) signaling from the posterior epiblast and posterior VE [33] is also required for PGC fate [27; 34]. WNT3 signaling upregulates the expression of transcription factors (TFs) including Brachyury (encoded by *T*), a mesodermal pro-

tein whose expression is required for PGC induction to occur [34]. BMP antagonists such as CER1 (Cerberus 1) and LEFTY1 (Left-right determination factor 1) and WNT3-agonist DKK1 (Dickkopf-related protein 1) are expressed from the anterior VE (AVE) [27]. The very small number of cells which receive the correct dosage of each activating signal and avoid the inhibitory signals will begin to upregulate PGC marker and key regulator BLIMP1 (B-lymphocyte-induced maturation protein 1; encoded by *Prdm1*) [26; 34] and emerge as the precursor PGCs [19].

By E7.25, the BLIMP1-positive cells also express AP2 $\gamma$  (encoded by *Tfap2c*) [35], and PRDM14 [36], and are considered founder PGCs [19]. These three proteins represent a “tripartite transcription factor” network which initiates the PGC gene program [37]. By E7.5, 30-40 PGCs are detectable in the embryo [38] and begin to express additional PGC marker genes including *Nanos3* (Nanos C2HC-type zinc finger 3), *Alpl* (Alkaline phosphatase liver/bone/kidney), and *Dppa3* (Developmental pluripotency associated 3) [39; 40], and will also begin to migrate towards the genital ridge, a somatic tissue which will ultimately form the testes or ovaries [41]. E7.5 PGCs are mitotically arrested at the growth 2 (G2) phase of the cell cycle, however they begin to rapidly proliferate from E8.5 until E12.5 [42]. By E10.5, the PGCs reach the genital ridge, and subsequently colonize and proliferate within the tissue [43].

A defining characteristic of PGCs is the process of epigenetic reorganization they undergo, which reduces the total methylation levels to the lowest observed during the lifetime of the organism [13] and leads to genome-wide alterations of histone modifications [44]. At the time of their induction, PGCs have highly methylated genomes, in which  $\sim 70\%$  of cytosines in CpG sequences are methylated [45]. Histone modifications undergo global remodeling, including the reduction of transcriptionally repressive H3K9me2 (Histone 3 lysine 9 dimethylation) marks from E8.0 onwards [44], while repressive H3K27me3 (Histone 3 lysine 27 trimethylation) marks and transcriptionally activatory H2A/H4R3me2 (Symmetrical dimethylation of arginine 3 on histone 2A and histone 4) marks are increased [46]. By E9.5, the percentage of total DNA methylation in the PGCs drops to  $\sim 30\%$  [45], indicating that the process of global epigenetic hypomethylation is underway. By E11.5, only 5% of the genome remains methylated [47], and demethylation will proceed even further until sex-specific development can begin at E12.5 [13]. In parallel to global DNA demethylation, female PGCs begin to repress long non-coding RNA (lncRNA) *Xist*

(Xi-specific transcript) and reactivate their second X chromosome [48], a process which is completed by E14.5 [48].

## 1.2.2 Key PGC factors

### 1.2.2.1 “Tripartite transcription factor network”

In 2013, Magnúsdóttir *et al.* characterized the three key PGC factors of AP2 $\gamma$ , BLIMP1, and PRDM14 in P19 embryonal carcinoma (P19) cells, a cell line derived from the E7.5 epiblast [37]. In this seminal work, they performed chromatin immunoprecipitation (ChIP) to identify where AP2 $\gamma$  and BLIMP1 bound in the P19 genome. This work broadly demonstrated that the three transcription factors cooperatively induced genes such as *Nanos3*, and also were bound to each other’s transcriptional regulatory sites, *i.e.* PRDM14 and AP2 $\gamma$  cooperatively were bound to the *Prdm1* regulatory site, while BLIMP1 bound to the *Tfap2c* regulatory site. Repressed somatic and mesodermal genes were often co-bound by all three factors, while activated PGC genes were often bound by PRDM14 and AP2 $\gamma$ , validating the role of BLIMP1 as a transcriptional repressor and demonstrating the high level of coordination between all three factors. It has since been accepted that AP2 $\gamma$ , BLIMP1, and PRDM14 are the three key factors of mouse PGC specification. These factors will thus be described in detail on the following pages.

### 1.2.2.2 AP2 $\gamma$ (*Tfap2c*)

AP2 $\gamma$  expression is directly induced by BLIMP1 in the small population of founder PGCs that arises on E7.5, and is also directly regulated by PRDM14 [37]. As one of the three key PGC factors, AP2 $\gamma$  is required for correct PGC fate in the mouse: conditional knockout (KO) of AP2 $\gamma$  specifically from SOX2-positive or TNAP- (Tissue-nonspecific alkaline phosphatase; encoded by *Alpl*) positive PGCs resulted in phenotypically normal but sterile pups, and the KO animals had small testes and ovaries relative to wild type (WT) animals [49; 35]. In the TNAP conditional KO model, PGCs could be identified by E8.0 in the KO embryos, however there were fewer PGCs than in WT controls and all TNAP-positive cells were lost by E12.5 [35]. AP2 $\gamma$  was also knocked-out in primordial germ cell-like cells (PGCLCs; Section 1.2.3), an *in vitro* recapitulation of PGCs, and microarray analysis of the transcriptomes of the KO cells demonstrated the downregulation of key PGC factors and upregulation of somatic genes [50]. AP2 $\gamma$  is also a human PGC factor [51] and has been identified as a putative pioneer factor in human PGCLCs, actively regulating an OCT4 enhancer

required for PGCLC fate [52]. AP2 $\gamma$  has additionally been identified as a PGC factor in rabbits [53], dogs [54], pigs [55], cynomolgus primates [56], and cows [57]. It appears that the role of AP2 $\gamma$  as a TF in germ cells is only embryonic; conditional KO of AP2 $\gamma$  in mouse pups three-to-four weeks after birth had no effect upon fertility [49].

### 1.2.2.3 BLIMP1 (*Prdm1*)

BLIMP1 has largely been regarded as a master regulator of PGC fate, as it is the first of the three key PGC genes activated in precursor PGCs; WNT3 specifically leads to the upregulation of *T*, which directly results in BLIMP1 and PRDM14 expression [34]. BLIMP1 was originally identified as a transcriptional repressor necessary to drive the differentiation of immunoglobulin-secreting plasma cells from B-cells [58], and was first characterized in the context of PGCs by Ohinata *et al.* in 2005 [26]. They found that BLIMP1-KO mouse embryos develop normally, however result in sterile adult mice [26]. The BLIMP1-KO PGCs failed to perform their characteristic migration from the site of induction to the genital ridge, and also failed to proliferate [26]. Moreover, BLIMP1-KO PGCs fail to repress somatic genes [40]. Interestingly, BLIMP1-KO and AP2 $\gamma$ -KO PGCs are phenotypically nearly identical [35]. BLIMP1 has also been identified as a human PGC factor [59; 60]

### 1.2.2.4 PRDM14 (*Prdm14*)

PRDM14 is a naïve pluripotency factor first detected heterogeneously in the four-cell stage embryo [5] which stabilizes the pluripotent state by repressing *de novo* methylation and actively recruiting ten-eleven translocation (TET) demethylating enzymes [61]. PRDM14 represses transcription of extraembryonic endoderm genes and upregulates stem-cell self renewal genes [62]. The expression of PRDM14 is induced by *T* by E7.25 and continues until E14.5 in the mouse [36]. PRDM14-KO mice appear phenotypically normal, however are born sterile; all TNAP-positive PGCs are lost by E12.5 [36]. E8.5 PRDM14-KO PGCs downregulate PGC markers and upregulate somatic and mesodermal markers [63], and PRDM14-KO PGCs also exhibit substantially higher DNA methylation than WT PGCs [64]. PRDM14 is also required for human PGC induction, where it regulates a different set of genomic targets than in the mouse, indicating evolutionary divergence [60].

### 1.2.2.5 Overlapping roles of PGC and naïve pluripotency markers

Induction of PGCs results in global DNA hypomethylation as well as the upregulation of naïve pluripotency markers, including *Dppa3*, *Nanog*, *Pou5f1*, *Prdm14*, and *Sox2*. It has been hypothesized that pluripotency factors help to stabilize the genome during epigenetic reprogramming in the PGCs, especially given that factors including *Dppa3* and *Prdm14* are implicated in maintaining demethylated chromatin [65]. Overall, the pluripotency factor expression of PGCs is incompletely understood and is still a topic of active research.

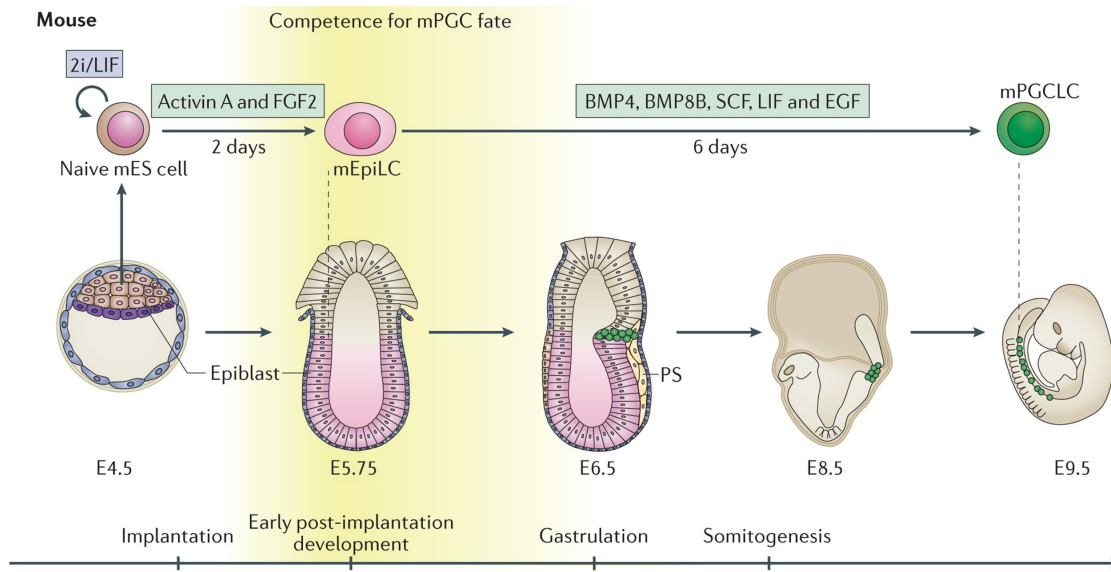
### 1.2.3 *In vitro* recapitulation of PGC induction

It is technically and ethically challenging to study *in vivo* mouse PGCs, as a population of only 30-40 PGCs is available on E7.5 [38], and isolation of PGCs can only be performed invasively and results in lethality for the embryo and mother. Thus, methods for *in vitro* PGC recapitulation have been studied not only to better understand how PGC specification may be induced, but also to ethically generate sufficient material for downstream analysis.

Currently, the gold standard method of *in vitro* PGC recapitulation is the production of PGC-like cells (PGCLCs) following the protocol published by Hayashi *et al.* in 2011 (Figure 4) [21]. In this protocol, ESCs are induced to EpiLCs, a previously described (Section 1.1.2) cell type comparable to cells of the E5.5-5.75 post-implantation epiblast. EpiLC induction occurs for two days in a defined medium containing activin A, bFGF, and KSR. EpiLCs are then induced towards PGCLC fate by forming aggregates of cells termed embryoid bodies (EBs) in anti-adherent plates in medium containing the cytokines BMP4, BMP8B, EGF (Epidermal growth factor), SCF (Stem cell factor), and LIF. The resulting EBs may be matured for 2-6 days to produce PGCLCs, which by day 6 of maturation transcriptionally and epigenetically resemble migratory PGCs of the E9.5 post-implantation embryo [24]. Male and female gametes which are derived from PGCLCs can be used to generate healthy offspring [21; 66].

It is also possible to generate PGCs by inducing ectopic gene expression. This approach has been shown to be successful at generating PGCs when ectopically expressing all three key PGC factors (AP2 $\gamma$ , PRDM1, PRDM14) in EpiLCs at once [67], or the individual protein PRDM14 in EpiLCs [67], NANOG in EpiLCs [68], or

LIN28 in ESC-based EBs [69].



**Figure 4: PGCLC induction and comparison to *in vivo* murine tissue.** Induction of PGCLCs begins with the culture of mouse ESCs (mES) in 2i/LIF (2i+LIF) medium. These cells may be induced directly towards mouse EpiLC (mEpiLC) fate for approximately two days in defined medium containing activin A and f FGF2 (Fibroblast growth factor 2; also known as bFGF). The resultant EpiLCs are comparable to cells of the E5.5-5.75 post-implantation epiblast. The EpiLCs may be directly induced towards mouse PGCLC (mPGCLC) fate by aggregating them as EBs in medium containing the cytokines BMP4, BMP8B, SCF, LIF, and EGF. After six days, they will resemble E9.5 migratory PGCLCs both transcriptionally and epigenetically. Modified from Tang *et al.*, 2016 [24].

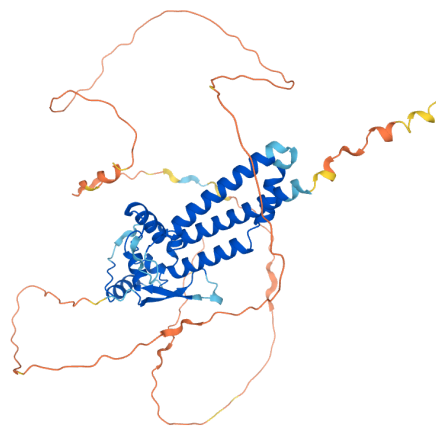
### 1.3 Transcription factor AP2 $\gamma$

Transcription factor AP2 $\gamma$  is the focus of this dissertation. AP2 $\gamma$  is not only a key factor for PGC fate, but also acts upstream of CDX2 during TE induction [70]. The implication of this finding is that AP2 $\gamma$  plays a deciding role in two of the earliest cell-fate decisions in the embryo: embryonic survival through placental development as well as maintenance of fertility through PGC specification. In literature, AP2 $\gamma$  is commonly described as a placental factor or PGC factor, but what is remarkable about the protein is that it appears to be expressed in a broad range of tissues, including the epidermis [71; 72; 73], retina [74], mammary gland [75; 76], embryonic neural crest [77], and adult hippocampus [78], and is also implicated in multiple forms of cancer [79]. To better understand the scope of this work, it is therefore necessary to describe what is already known about AP2 $\gamma$ .

### 1.3.1 AP2 $\gamma$ overview

The AP2 transcription factor family comprises five members -  $\alpha$  [80; 81],  $\beta$  [82; 83],  $\gamma$  [71; 84; 83],  $\delta$  [85], and  $\epsilon$  [86; 87]. All members of the AP2 TF family are expressed in the embryo, and all members except AP2 $\delta$  are widely-expressed in the pre-implantation embryo [88]. The factor of interest to this work is AP2 $\gamma$  (Figure 5). AP2 $\gamma$  is a TF which acts as both a repressor and activator of transcription [37], and was discovered in 1996 by Chazaud *et al.* as a retinoic acid (RA) responsive mRNA in P19 cells [84]. The protein was originally termed AP-2.2, as it exhibited 65% sequence similarity with already-characterized AP2 family proteins in the mouse and human, and contained a 96-nucleotide conserved AP2 binding domain [71], which bound to the only AP2 consensus sequence identified at the time, 5'-GCCN<sub>3</sub>GGC-3' [80]. AP-2.2 was identified to be primarily expressed in the male and female sex organs and squamous epithelia [71]. The protein was later renamed AP2 $\gamma$  to comply with the naming scheme of previously discovered AP2 $\alpha$  and AP2 $\beta$ . AP2 $\gamma$  is translated from the *Tfap2c* mRNA transcript, although in older literature it may be referred to as *Tcfap2c*.

AP2 $\gamma$  has two recognized isoforms (Figure 6A and B) [91]; isoform 1 is expressed as a 449 amino-acid (aa) protein [92] while isoform 2 is expressed as a 488 aa protein [93]. A third 513 aa isoform is predicted by the UniProt database [94]. The two recognized isoforms differ only in their transcriptional start sites (TSSs) and first exon, and exhibit 100% sequence homology between exons two through seven [91] (Figure 5C). The translated protein isoforms share several features [91; 92; 93]: two disordered domains, a proline-rich (PPxY) motif, a conserved transcription factor AP-2 domain which harbors a dimerization/helix-span-helix domain, and two predicted phosphoserine sites (Figure 5D and E). The PPxY motif is of interest as it binds to the *Wwox*



**Figure 5: Predicted protein structure of AP2 $\gamma$ .** The predicted protein structure of AP2 $\gamma$  isoform 1 generated by AlphaFold Monomer v2.0 [89; 90]. The per-residue confidence score is indicated by color: high confidence - confidence - low - very low is indicated by dark blue - light blue - yellow - orange.





tein [92]. The attachment of a SUMO protein to another protein is a process termed SUMOylation, and has many described potential functions in the cell, including facilitation of protein-protein interaction, disruption of protein-protein interaction, DNA damage response, proliferation, and apoptosis, among many others [99]. It has been previously shown that SUMOylation of AP2 $\gamma$  acts as a repressor of transcriptional activity [100].

AP2 $\gamma$  is commonly associated with its fellow AP2 family member AP2 $\alpha$ . AP2 $\gamma$  exhibits 52% aa sequence similarity with AP2 $\alpha$  [101], they both exhibit a SUMOylated lysine on their N-termini of some isoforms [100], they share the conserved AP-2 domain [97], and they are both putative pioneer factors [52; 73; 77; 102; 103] - transcription factors which “open up” previously inaccessible chromatin for other TFs and adapter proteins [104]. They are often co-expressed in the same tissues, and thus the question of whether the two proteins are redundant has often been raised. Genetic redundancy describes two or more genes with overlapping functions; if one gene is knocked out, the other(s) are able to functionally compensate for the loss [105]. The potential redundancy of AP2 $\alpha$  and AP2 $\gamma$  has been previously investigated, with examples including: AP2 $\gamma$  function was proposed to compensate for AP2 $\alpha$ -KO in the epidermis of mice [106], embryonic dual AP2 $\gamma$ - and AP2 $\alpha$ -KO results in an earlier embryonic lethality than AP2 $\gamma$ -KO alone [88], and functional redundancy of AP2 $\gamma$  and AP2 $\alpha$  was demonstrated in neural crest induction in the zebrafish [107]. The potential redundancy of AP2 $\alpha/\gamma$  in other tissues and contexts is still an intriguing open question.

AP2 $\gamma$  is widely implicated in a number of human cancers [79] including breast cancer [108; 109; 110; 111], testicular cancer [112], germ cell tumors [112; 113], non-small cell lung cancer [114; 115; 116], pancreatic ductal adenocarcinoma [117], bladder cancer [118; 119], colorectal cancer [120], and neuroblastoma [121]. Intriguingly, *Tfap2c* has been identified as both a cancer repressor [108; 109; 115; 118] and as a driver of tumorigenicity [110; 111; 112; 114; 116; 117; 119; 120; 121]. The published data describing the effect of AP2 $\gamma$  on the same types of cancer can be conflicting - Kang *et al.* describe AP2 $\gamma$  increasing the aggressiveness of non-small cell lung cancer [114], while Chang *et al.* describe micro-RNA (miRNA)-mediated knockdown (KD) of AP2 $\gamma$  resulting in a more aggressive form of the cancer [115]. In another example, AP2 $\gamma$  has been shown to repress the expression of the cancer-associated *Myc* (myelocytomatosis oncogene, also known as *c-Myc*) gene in somatic cells [122],

while in P19 cells AP2 $\gamma$  and PRDM14 synergistically induced *Myc* expression [37]. It is therefore likely that the activity and potency of AP2 $\gamma$  is extremely context-dependent.

### 1.3.2 AP2 $\gamma$ in early embryonic development

*Tfap2c* maternal mRNA is already present in the egg at the moment of fertilization, and thus AP2 $\gamma$  is present in the early embryo prior to ZGA [88]. However, KO of only the maternal *Tfap2c* transcript has no affect upon embryonic development [88]. At the onset of ZGA, *Tfap2c* is immediately expressed in the two-cell stage embryo and throughout pre-implantation development [88]. Following TE induction by E3.5, equal amounts of AP2 $\gamma$  may be found in the nuclei of cells in both the ICM and TE [88], however *Tfap2c* is downregulated in the epiblasts of late blastocysts, and after implantation is primarily expressed in all TE derivates except the syncytiotrophoblast cells [123]. As previously mentioned, dual KO of both *Tfap2c* and *Tfap2a* results in embryonic lethality between E3.5-E7.5 [88], which is earlier than the embryonic lethality occurring by approximately E8.5 [124] in a single AP2 $\gamma$ -KO line, suggesting a potentially redundant role in the blastocyst between AP2 $\alpha$  and AP2 $\gamma$  [88].

One of the earliest papers to investigate the role of AP2 $\gamma$  in TE induction was published by Auman *et al.* in 2002 [124]. They demonstrated that AP2 $\gamma$  is required for extraembryonic tissue establishment as well as embryonic survival, as ablation of exon 6 and part of exon 7 of *Tfap2c* resulted in embryonic death by E8.5-10.5 due to defects in extraembryonic tissues. Prior to E8.5, embryos with mutated AP2 $\gamma$  were smaller than their littermates and often did not exhibit a primitive streak. Tetraploid-diploid embryonic chimeras, in which *Tfap2c*-mutated cells contributed to the embryo proper but had WT extraembryonic tissues, developed normally and survived for up to one year after birth, indicating that AP2 $\gamma$  is dispensable for the development of the embryo proper. However, a TE-specific AP2 $\gamma$ -KO mouse resulted not only in TE failure and resultant embryonic death, but also developmental failure in the embryonic tissues. This work suggested that AP2 $\gamma$  plays an important regulatory role in embryonic development through its activity in the TE, and further demonstrated how extraembryonic signaling is a required component of embryonic development. It is to be noted that PGC specification is a good example of how extraembryonic signaling is required for some elements of embryonic development

(Section 1.2.1).

Once it had been established that AP2 $\gamma$  expression is required for correct TE induction, the focus of research shifted to better understand how AP2 $\gamma$  regulates TE induction. *In vitro*, induced ectopic expression of *Tfap2c* is sufficient to induce ESCs to trophoblast stem cell (TSC) fate, even in CDX2-KO cell lines [125], and these induced TSCs differentiated toward TGC fate when placed in differentiation medium [125]. *In vivo*, AP2 $\gamma$  has been shown not only to directly regulate TE factor CDX2 [124], but also to initiate its expression in the early embryo by acting as the upstream regulator of *Cdx2* in the two- to eight-cell stage embryo [70]. By the eight-cell stage, AP2 $\gamma$  begins to negatively regulate the hippo signaling pathway [70]; hippo signaling is a necessary pathway driving the first fate decision of the early embryo (see Section 1.1.1), thus AP2 $\gamma$  plays a vital role in ensuring the TE is correctly specified from the ICM.

The role of AP2 $\gamma$  in the development of the embryo was revisited by Choi *et al.* in 2012 [126]. In their study, all maternal and zygotic *Tfap2c* mRNA was knocked-down in one-cell stage embryos using RNA interference (RNAi). Only 13.73% of the KD embryos formed blastocysts, relative to 92.64% of siRNA controls which formed blastocysts, and it was observed that most of the *Tfap2c*-KD embryos failed to form a blastocoel cavity. Analysis of transcriptomic and ChIP-seq data revealed that AP2 $\gamma$  regulates a suite of genes required for tight-junction (TJ) assembly, cell polarity, and fluid accumulation. The authors state, “[t]o our knowledge, this is the first study to describe the function of a single TF controlling TJ assembly, fluid accumulation, and cell proliferation in early embryogenesis”, and suggest that the phenotypically normal blastocysts seen in prior AP2 $\gamma$ -KO studies [88; 124] did not account for the presence of the maternal *Tfap2c* mRNA. This work dramatically broadens the potential scope and import of AP2 $\gamma$  in early embryogenesis.

### 1.3.3 AP2 $\gamma$ in mammary development

#### 1.3.3.1 Mammary gland development in the mouse

Embryonic mammary development begins on E10.5 in the dermal mesenchyme of the mouse with the formation of the “milk line”, an ectodermal structure which forms five placodes by E11.5 [127]. Subsumption of the placodes by the mesenchyme by E13.5 generates the ten mammary buds, which sprout into the incipient fat pads on

E15.5 [128]. By E18.5, an early ductal tree composed of one primary duct with ten-to-fifteen secondary branches emerges per each bud, as does the rudimentary nipple sheath [128]. It is only after birth and at the onset of puberty that development resumes, and colonization of each mammary fat pad with mammary ducts is led by terminal end buds (TEBs), structures at the ends of ducts comprised of luminal-epithelium (luminal cells) and myo-epithelium (basal cells) [128]. A population of fetal mammary stem cells (fMaSCs) has been identified by E18.5, and adult MaSCs are resident in adult tissue [128]. Two important lineages of mammary progenitor cells arise from fMaSCs: luminal and basal cells. Luminal cells are the epithelial cells which line the inner lumen of mammary ducts and contains secretory alveoli, while basal cells comprise the outer layer of the mammary ducts [129].

### 1.3.3.2 AP2 $\gamma$ in mammary gland development

AP2 $\gamma$  was first implicated in the post-natal development of mammary tissue in mice in 2003 in a study by Jäger *et al.*, in which overexpression of AP2 $\gamma$  resulted in the hyperproliferation of the alveolar epithelium in pregnant mice [130]. It has since been found that AP2 $\gamma$ -KO results in impaired mammary gland branching during pre-pubertal development [131], and conditional KO demonstrates that AP2 $\gamma$  expression is limited to the luminal epithelial cells of the adult mammary gland [76]. Stable KD of AP2 $\gamma$  in human luminal breast cancer cells *in vitro* resulted in a transition of the cells towards basal fate [76]. By seven weeks post-birth, both AP2 $\gamma$  and AP2 $\gamma$  are detectable in basal and luminal mammary epithelial cells of the mouse, and AP2 $\gamma$  is detected in the mammary stem cells of post-natal mice (MaSCs) [75].

## 1.4 *Cis*-regulatory elements

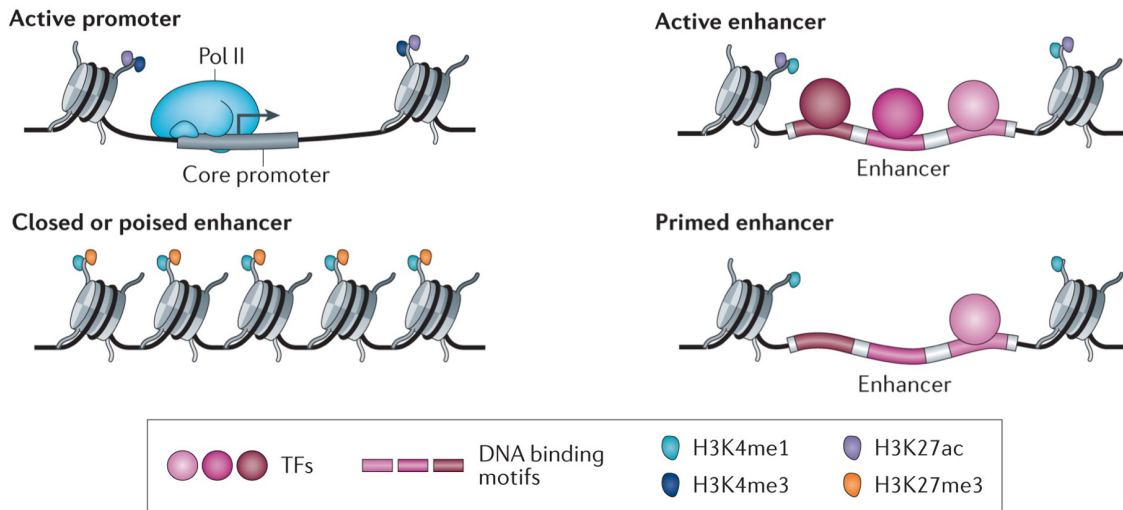
According to the Mouse Genome Informatics (MGI) group of the Jackson Laboratory, there are 23,151 currently accepted protein-coding genes in the mouse [132]. Every gene is not expressed uniformly or continually; the correct combination and expression of genes at specific time points is required for normal development and homeostasis. The coordination of combinatorial gene expression is thus a topic of great interest; by better understanding how genes are expressed and what is responsible for their expression, we can better unravel the intricacies of the cell and also potentially identify regulatory origins of human disease. Enhancers are *cis*-regulatory elements which upregulate the transcription of target genes [133]. In the mouse, approximately 80% of enhancer activity is tissue-specific [134], and it is believed that

enhancers actively coordinate the expression of genes required for cell fate transition. Thus, one aim of this work was to identify and functionally characterize enhancers acting on *Tfap2c*, to better understand the upstream regulation of the gene as well as to identify enhancers required for PGC fate. To establish the background of this project, current knowledge of regulatory elements must be explained.

### 1.4.1 Identifying and classifying CREs

*Trans*- and *cis*-regulatory elements are key regulators of gene expression. *Trans*-regulatory elements encode TFs; proteins which bind to DNA and recruit RNA polymerases or co-activator/repressor proteins and complexes in order to affect transcription [133]. *Cis*-regulatory elements (CREs) are non-coding DNA sequences which regulate the expression of target genes [135]. This is achieved through the binding of TFs and other adapter/modulator proteins to motif sequences enriched within CREs, which then allows the element to interact with a target CRE or gene and subsequently regulate transcription [135]. CREs broadly describe four classes of non-coding DNA elements: promoters, insulators, silencers, and enhancers [136]. Promoters are required for eukaryotic gene transcription and are located close to the TSS of a target gene [135]. Generally there is one promoter per gene, although under specific circumstances this may not be true [135]. Active promoters are nucleosome-free and may be indicated by the presence of flanking regions enriched with H3K27ac (Histone modifications histone 3 lysine 27 acetylation) and H3K4me3 marks (Histone 3 lysine 4 trimethylation; Figure 7) [137]. Insulators create physical barriers in the genome, thus preventing the formation of certain chromatin-chromatin interactions [136], and also stopping the spread of transcriptionally repressive heterochromatin marks [138]. Insulators may be indicated by enrichment of the transcription factor CCCTC-binding factor (CTCF) [138]. Silencers are transcriptionally repressive regulatory elements, and may be indicated by the histone modifications H4K20me1 (Histone 4 lysine 20 monomethylation) and H3K27me3 [136].

Unlike promoters, enhancers may regulate multiple genes, and are not required to be adjacent to the TSS to function; enhancers have been identified in the introns of others genes, or up to a megabase (Mb) away from their target gene [139]. Enhancers upregulate target genes through direct proximity with their promoters in a process known as looping: a “loop” of chromatin forms when proteins such as cohesin enable the TF-bound active enhancer to interact with its target promoter,



**Figure 7: Histone modifications indicate the activity status of enhancers and promoters.** Histone modifications allow for the discrimination of CREs based on activity. Nucleosome-free promoters with flanking regions enriched for H3K27ac and H3K4me3 are classified as active, and nucleosome-free enhancers in with flanking regions enriched for H3K4me1 and H3K27ac are also classified as active. Closed or poised enhancers may be identified by the presence of nucleosomes and the histone modifications H3K27me3 and H3K4me1, while a primed enhancer may be nucleosome-free and flanked with regions enriched for H3K4me1. RNA polymerase II (Pol II), transcription factors (TFs), and DNA binding motifs are indicated. Modified from Shlyueva *et al.*, 2014 [137].

resulting in recruitment of RNA polymerase II (pol II) and initiating transcription [137]. Enhancer-target interactions and subsequent chromatin loops may form extremely complex structures, and the mechanisms controlling and maintaining these structures represent a very active field of research [140]. Active enhancers are typically nucleosome-free, and are flanked by chromatin displaying histone modifications including H3K27ac and H3K4me1 (Histone 3 lysine 4 monomethylation; Figure 7) [137]. Poised enhancers are enhancers which share some but not all characteristics of active enhancers, and may exist in preparation for a subsequent cell state in which their activity is required [141], and are flanked by chromatin enriched with histone modification H3K4me1 [137]. Inactive closed or primed enhancers may be indicated by the histone modifications H3K4me1 and H3K27me3 [137].

Enhancers may be classified not only based on activity status, but also on behavior. Redundant enhancers (also known as “Shadow enhancers” in *D. melanogaster*) are enhancers which exhibit partial or completely overlapping functional activity [142]. It has been shown in the fruit fly that shadow enhancers are not only able to cooperatively upregulate a gene, but are also capable of repressing each other and resulting in transcriptional downregulation [142]. A 2018 study by Osterwalder *et al.* found

that developmental genes in the mouse generally had multiple, redundant distal enhancers while housekeeping genes often had none [143]. Another type of enhancer which may be classified based on behavior is the super-enhancer. Super-enhancers represent a small proportion of all enhancers, but are key players in development. Super-enhancers are larger than normal enhancers and may be identified bioinformatically by the presence of H3K27ac modifications and binding of the p300 and mediator complexes [144]. More importantly, super-enhancers are enriched with lineage-specific TFs, and are believed to be necessary for cell-state transitions or maintenance [144].

Putative enhancers have been characterized by the The Encyclopedia of DNA Elements (ENCODE) project through the integration of next generation sequencing (NGS) techniques, including: DNase treatment followed by sequencing assay (DNase-seq), assay for transposase-accessible chromatin using sequencing (ATAC-seq), and ChIP-seq [145]. DNase-seq and ATAC-seq were performed to allow for the identification of accessible chromatin, which is characteristic of active enhancers, while ChIP-seq was performed to localize histone modifications, histone variants, and histone-modifying proteins to identify characteristic indications of enhancers as well as their activity status [145]. By generating these data-sets (as well as many others) and integrating the information *in silico*, 339,815 putative enhancers have been identified in the mouse [145], outnumbering protein-coding genes by almost 15:1. Enhancers may also be predicted based upon the presence of TF binding sites, transcription of the enhancer as enhancer RNA (eRNA), or sequence similarity to evolutionarily conserved enhancers in other species [146]. A method to identify potential enhancer interactions is Hi-C chromatin conformation capture (Hi-C), which allows one to crosslink the three-dimensional (3D) genome, digest the genome, and ligate the fragments, resulting in linear sequences comprised of two chromatin contact-sites [147]. The resulting libraries may be enriched for specific bait sequences, *e.g.* a set of promoter sequences, prior to sequencing in order to identify chromatin interacting with specific sequences of interest [148]. This technique was originally termed promoter-capture Hi-C [148], but can be used for any set of baits, including enhancer sequences, silencer sequences, or any custom list of potential genomic interactions. It should be noted that capture Hi-C does not necessarily demonstrate functional *in vivo* interactions [146].



## 1.4.2 Functional testing of enhancers with CRISPR/Cas9

The identification of putative enhancers does not prove *in vivo* functionality. One method to test the functionality of an enhancer is to introduce a mutation into a TF binding site within the enhancer and observe the resultant phenotype. This may be accomplished using clustered regularly interspaced short palindromic repeats/CRISPR associated protein 9 (CRISPR/Cas9) [146].

CRISPR/Cas9 is a molecular technique modified from a prokaryotic immune strategy, wherein a bacterial or archaeal Cas enzyme digests pathogenic DNA adjacent to specific 2-4 nucleotide-long sequences termed protospacer adjacent motifs (PAMs) [149]. The digested DNA is stored in the genome in a CRISPR array as a “spacer” sequence [149]. The spacer sequence will be transcribed and processed, and the mature sequence will be retained in the cell as a CRISPR RNA (crRNA) [149]. The crRNA forms a complex with a *trans*-activating CRISPR RNA (tracrRNA), and the RNA complex guides the Cas9 nuclease to pathogenic DNA and thus targets it for destruction [150]. *In vitro*, a single guide RNA (gRNA) substitutes for the crRNA/tracrRNA complex, and can be used to guide the Cas9 to a specific ~20 nucleotide (nt) genomic region adjacent to a PAM site to introduce a double strand break (DSB) [150]. Generally, Cas9 deriving from *Streptococcus pyogenes* (*S. pyogenes*) is used, and thus the *S. pyogenes*-specific 5'-NGG-3' PAM sequence [150] must be considered when designing gRNAs. The DSB will be repaired intracellularly in one of two ways: via non-homologous end joining (NHEJ) or homology directed repair (HDR) [150]. NHEJ often results in the addition or deletion of several nucleotides following DSB repair, and thus is used to introduce frame-shift mutations into coding sequences to KO a target gene [150]. HDR requires a repair template, which provides the cellular DNA repair machinery a sequence to insert at the mutagenized locus [150].

Enhancers are non-coding [135], so NHEJ may be used to disrupt the binding motif of a specific TF within a transcription factor, but is not sufficient to functionally KO an entire enhancer. To excise the entire enhancer, two gRNAs are generally employed; one targeting the 5' border of the enhancer, and the other targeting the 3' border. This method, termed multiplex CRISPR gene editing [151], allows for a targeted region to be completely excised from the genome. This approach was employed in the previously discussed work by Osterwalder *et al.* (Section 1.4.1) [143] and has been routinely and successfully performed elsewhere [146; 152].

### 1.4.3 CRISPR/Cas9 genetic screens

KO screens are commonly performed to identify previously overlooked genes in a given cellular context. Several commercial gRNA libraries exist which provide multiple exon-targeting gRNAs per human or mouse gene, including the genome-scale CRISPR knockout (GeCKO) library [153], the improved genome-wide knockout CRISPR library v2 [154], the CRISPR knockout pooled library (Gouda/Brunello) [155], and many others. These gRNAs can be used in a pooled lentiviral screen: in this approach, one gRNA on average is introduced per cell and thus results in a pooled population of cells containing unique mutations [146]. In a dropout screen, cells which are lost from the pool due to decreased proliferation or survival may be identified, as their specific genome-integrated gRNA sequences will be depleted relative to the overall population of gRNAs [156]. This approach is commonly performed to screen protein-coding genes, as the gRNAs target coding exons and thus may introduce a frameshift and KO the protein, resulting potentially in a dramatic phenotype and loss of cellular fitness.

It is considerably more challenging to screen putative enhancer elements, as enhancers are non-coding and thus the introduction of a frameshift is not possible, and the mutagenization of one region within the enhancer may not affect the overall function of the enhancer [146]. Thus, in order to use CRISPR/Cas9 and directed mutagenesis to screen the function of several enhancers, creative approaches must be employed. Two screens of particular interest to this work were published in 2016 by Sanjana *et al.* [157] and in 2020 by Canver *et al.* [158], respectively. In the Sanjana *et al.* publication, a tiling screen approach is described in which 100 kilobase (kb) regions up- and downstream of target putative oncogenes were tiled with gRNAs and negatively selected by strong anti-cancer drugs. This approach successfully identified regulatory elements which, when mutagenized, allowed the cancer cells to survive [157]. In the Canver *et al.* publication<sup>2</sup>, all putative regulatory elements acting in ESCs were mapped using ATAC-seq and targeted for saturating mutagenesis by tiling the regions with as many gRNAs as the number of PAM sites permitted [158]. The 12 kb regulatory region surrounding the gene *Pou5f1* was also targeted for saturating mutagenesis. This screen allowed for the identification of regulatory elements acting on *Pou5f1*, and the approach was termed *cis*- and *trans*-regulatory

---

<sup>2</sup>In contravention to accepted nomenclature, this paper refers to distal enhancers as *trans*-regulatory elements, and proximal enhancers as *cis*-regulatory elements

elements scanning through saturating mutagenesis and sequencing (ctSCAN-SMS).

## 1.5 Scope and aims of work

AP2 $\gamma$  is required for the correct formation of the blastocyst, the induction of TE fate, and the specification of PGCs. In its absence the embryo cannot survive, and even conditional KO models demonstrate dramatic and permanent consequences for development. It is also the least characterized key PGC factor within the context of PGC specification. Thus, my project aimed to characterize: 1) the regulatory activity of *Tfap2c* during PGC induction, and 2) the regulatory elements acting on *Tfap2c* during PGC induction. By better understanding the upstream regulators acting on *Tfap2c* as well as the genomic targets and consequences of *Tfap2c* activity, we can better understand how PGCs are correctly specified, and potentially identify regulators or regulatory targets which may be implicated in infertility.

My work aimed to improve our understanding of the role AP2 $\gamma$  plays in PGC specification and differentiation by first knocking-out AP2 $\gamma$ -KO in ESCs and examining the effect upon the transcriptome and chromatin accessibility of ESCs, EpiLCs, and cells induced towards PGCLC fate via RNA-seq and ATAC-seq. These NGS-based approaches allowed for the assessment of AP2 $\gamma$ -KO as a transcriptional regulator and potential pioneer factor. Ultimately, I was able to identify misregulated genes and differentially-accessible chromatin present in the AP2 $\gamma$ -KO cells, and demonstrates that AP2 $\gamma$  is required for correct PGCLC induction in mouse cells. Further, this work raises questions about the potentially overlapping but not redundant role of AP2 $\gamma$  and AP2 $\alpha$  in the PGCLCs.

To identify CREs acting on *Tfap2c*, I employed a CRISPR/Cas9-based lentiviral screen approach by tiling the topologically active domain (TAD) of the gene *Tfap2c* with gRNAs. TADs are defined, self-contained units of chromatin which may exist to facilitate and confine regulatory element-target gene interactions within a certain region [159]. Thus, the *Tfap2c* TAD was targeted as it is believed that most regulatory elements acting on a target gene are present within the same TAD as the gene [160]. This is the first known tiling screen of an entire TAD, and was performed not only to identify putative *Tfap2c* CREs, but also to test the functionality of this CRISPR-based screen approach. Although I was not able to identify putative CREs acting solely on *Tfap2c*, the information collected during the screen may be used to

identify and characterize potential cell type-specific CREs, as well as to better design future screens to identify regulatory elements.

# Materials and methods

All materials, software, and equipment used in this thesis may be found in the appendix. Additional buffers not listed in the materials and methods may also be found in the appendix. All chromosomal coordinates provided are from the GRCm38 (mm10) mouse genome assembly [161]. All next-generation sequencing occurred at the Sequencing Core Facility of the Max Planck Institute for Molecular Genetics.

## 2.1 Cell culture

All cell lines used in this work are listed in Appendix 7.4.

### 2.1.1 ESC culture

Murine ESCs were cultured continuously in 2i+LIF medium [14] (Appendix 7.3.1) in 12-well Nunc<sup>TM</sup> cell-culture multiwell plates (Thermo Scientific) in a 37°C, 5% CO<sub>2</sub> HERAcell<sup>TM</sup> VIOS 160i humidified incubator (Thermo Scientific). Prior to passaging, plates were coated with 1:60 human plasma fibronectin (Sigma-Aldrich) in 1x phosphate-buffered saline (PBS; Gibco) for at least thirty minutes at 37°C. To passage, approximately 70-80% confluent cells were washed with 1 ml PBS and incubated with 100  $\mu$ l TrypLE<sup>TM</sup> Express (Gibco) for 4 minutes at 37°C in a humidified incubator. 1 ml of passaging medium (10% FBS (Gibco), 100 U/ml Penicillin/Streptomycin (P/S; Gibco), 2 mM L-Glutamine (Gibco) in GMEM (Gibco)) was then added to each well, and the cells were gently pipetted up and down to dissociate. The cell suspension was then transferred to a 15 ml tube, and centrifuged at 1,200 xg for 4 minutes at room temperature (RT). The supernatant was aspirated and the pellet was gently resuspended in 2i+LIF medium. The cell concentration could then be counted using the Countess II automated cell counter (Invitrogen), and an appropriate volume of cells was subsequently diluted into 1 ml of 2i+LIF medium on the pre-coated plate. To maintain the ESCs in culture, approximately  $1 \times 10^4$  cells were plated per well, the medium was replaced every other day, and the cells were passaged every 3-4 days as needed. Prior to transfection or EpiLC induction,  $1 \times 10^5$  cells were plated per well.

### 2.1.2 EpiLC and PGCLC induction

EpiLC and PGCLC inductions were performed according to the protocol developed by Hayashi *et al.* [21]: First, ESCs were dissociated as previously described (Section 2.1.1). After the cells were suspended in passaging medium, centrifuged, and

resuspended in 1 ml 2i+LIF, the concentration of cells was counted using the automated cell counter. An appropriate volume of cells was then pipetted into a fresh 15 ml tube and centrifuged for 4 minutes at 1,200 xg, RT. The pellet was gently resuspended in EpiLC medium (1% Knockout Replacement Serum (KSR; Gibco), 20 ng/ml Activin A (Wellcome-MRC Cambridge Stem Cell Institute), 12 ng/ml FGF2 (Wellcome-MRC Cambridge Stem Cell Institute) in N2B27 medium) such that a concentration of  $1 \times 10^5$  cells/ml medium was reached. 1 ml of cell suspension was then plated per well of a 12-well plate, which had been pre-coated with 1:60 fibronectin in PBS for at least one hour at 37°C. The medium of the cells was replaced after 24 hours, and after 42 hours in a humidified incubator, EpiLC induction was complete and cells could be further induced towards PGCLC fate or harvested for downstream analysis.

To induce PGCLCs, EpiLCs were washed with PBS and incubated with 100  $\mu$ l TrypLE™ Express per well for four minutes at 37°C in a humidified incubator. The cells were then resuspended in 1 ml passaging medium and pipetted into a 15 ml tube, and the suspension was centrifuged for four minutes at 1,200 xg, RT. The pellet was then gently resuspended in 1 ml GMEM with 15% KSR (GK15) medium (15% KSR, 1x non-essential amino acids (NEAA; Gibco), 1 mM Sodium-pyruvate (Sigma-Aldrich), 2 mM L-glutamine, 100 U/ml P/S, and 0.1 mM 2-Mercaptoethanol (Gibco) in GMEM). The concentration of cells was determined using the automated cell counter, and an appropriate volume of cells was transferred to a 1.5 ml microreaction tube and centrifuged at 1,200 xg for 4 minutes at RT. The supernatant was aspirated and the cells were resuspended in either unsupplemented GK15 medium (“-cyto”) or GK15 medium with cytokines (“+cyto”) comprising 500 ng/ml BMP4 (R&D Systems), 500 ng/ml BMP8a (R&D Systems), 10 ng/ml LIF (Wellcome-MRC Cambridge Stem Cell Institute), 100 ng/ml SCF (R&D Systems), and 50 ng/ml EGF (R&D Systems). EpiLCs placed in the -cyto medium would not induce towards PGCLC fate, while approximately 10-30% of EpiLCs placed in the +cyto medium would be correctly induced into PGCLCs.

Embryoid bodies (EBs) were then formed using one of two methods: 1) According to the Hayashi *et al.* protocol [21], EpiLCs were resuspended in +/-cyto GK15 medium to achieve a concentration of  $2 \times 10^4$  cells/ml. 100  $\mu$ l of this suspension was pipetted per well of a round-bottom, low-adherence Costar™ 96-well microtiter plate (Corning), and the EBs were allowed to grow in the humidified incubator for

2-6 days, at which point they could be collected or dissociated for further analysis. EBs were supplemented on day four of incubation with 50  $\mu\text{l}$  fresh -cyto GK15 per well, if they were to be grown for 5-6 days in total. 2) Following the Gruhn and Günesdogan protocol [162], EpiLCs were resuspended in +/-cyto GK15 medium to achieve a concentration of  $7.5 \times 10^5$  cells/ml, and 1 ml of the suspension was gently pipetted per well of a 6-well EZSPHERE™ Microwell plate (Asahi Glass Company). The cells were gently distributed by shaking the plate horizontally and vertically in the humidified incubator, and were not touched for 24 hours. The next day, 1 ml of fresh -cyto GK15 medium was gently added per well, and the plate was again gently shaken. EBs could be collected after 2-6 days total incubation time. From days 3-6 of incubation, 1 ml of medium per well was gently removed with a 1 ml pipette, a fresh ml of -cyto GK15 was gently supplemented, and the plate was gently shaken horizontally and vertically to distribute the EBs throughout the well.

### 2.1.3 Transfection of ESCs

The transfection protocol for ESCs was modified from the Lipofectamine™ 2000 manufacturer's protocol [163] and is described in brief: The day before transfection,  $1 \times 10^5$  ESCs were plated per well of a 12-well plate pre-coated with fibronectin as previously described (Section 2.1.1). The next day, 1  $\mu\text{l}$  of Lipofectamine™ 2000 (Invitrogen) per number of total plasmids required for the transfection was mixed with 50  $\mu\text{l}$  of Opti-MEM™ I reduced serum medium (Gibco). In parallel, 500 ng of each plasmid was mixed with 50  $\mu\text{l}$  of Opti-MEM™ medium. Both solutions were incubated at RT for five minutes, and were then subsequently mixed and incubated at RT for twenty minutes. 100  $\mu\text{l}$  of the subsequent solution was combined with 900  $\mu\text{l}$  of transfection medium (10% FBS, 100 U/ml P/S, and 20 ng/ml LIF in GMEM) [164] and gently pipetted into the well of plated cells whose medium had just been aspirated. The transfection reaction proceeded for three hours at 37°C, 5% CO<sub>2</sub> in a humidified incubator. The transfection medium was subsequently aspirated and replaced with fresh 2i+LIF medium containing selection antibiotics. Puromycin (Gibco) was used at a concentration of 1.2  $\mu\text{g}/\text{ml}$  for two days of selection, or hygromycin was used at a concentration of 200  $\mu\text{g}/\text{ml}$  for five days of selection. The selection medium was replaced with fresh antibiotic-containing medium each day until the selection period was complete. The pooled transfected cell lines could then be characterized directly, or sorted for clonal line generation and characterization using fluorescence-activated cell sorting (FACS; Section 2.3).

### 2.1.4 293T-HEK Culture

Human embryonic kidney 293T cells (HEK cells) were cultured in D10 medium (10% FBS, 1x GlutaMAX™ supplement (Gibco), and 100 U/ml P/S in DMEM/F12 (Gibco)) [156]. HEK cells were passaged once they had reached a confluency of 80-90% according to the protocol described in Section 2.1.1, however when passaging HEK cells the dissociation incubation period following the addition of TrypLE™ was increased to five minutes. Following resuspension in passaging medium and centrifugation, the HEK pellet was resuspended in fresh D10 medium, and passaged at a ratio of 1:4 onto 10 cm Nunc™ EasYDish™ dishes (Thermo Scientific) or 225 cm<sup>2</sup> Nunc™ EasYFlask™ cell culture flasks (Thermo Scientific) pre-coated with 1:5 0.1% Poly-L-lysine (Sigma) in PBS for 10 minutes at RT.

## 2.2 Generation of genetically-modified ESC lines

### 2.2.1 Single-gRNA CRISPR/Cas9-based KO lines

CRISPR/Cas9 gene editing requires a guide RNA (gRNA) to direct the Cas9 enzyme to a specific genomic locus and produce a double-strand break (DSB) [150]. In this work, single gRNAs of 20 nucleotides (nt) in length targeting exon 2 of the gene *Tfap2c* were designed using the Graphical User Interface for DNA Editing Screens (GUIDES) tool of the Neville Sanjana lab at New York University (NYU) [165]. Once the gRNAs had been designed, DNA oligos comprising the gRNA sequence were ordered according to the Feng Zhang lab at the Massachusetts Institute of Technology (MIT)'s target sequence cloning protocol [156], which is depicted below, and where twenty Ns represent the unique gRNA sequence:

Forward oligo: 5' - CACC(G)NNNNNNNNNNNNNNNNNNNNNNNN - 3'

Reverse oligo: 3' - (C)NNNNNNNNNNNNNNNNNNNNNNNNCAAA - 5'

Designed oligos were ordered from Integrated DNA Technologies (IDT), and their sequences may be found in Appendix 7.10.1.1). The lyophilized oligos were then reconstituted with UltraPure™ DNase/RNase-Free distilled water (dH<sub>2</sub>O; Invitrogen) to a concentration of 100 μM. Ligation of the forward and reverse oligos was performed according to the Ran *et al.* protocol [166], in short: a solution containing 45 μM of each oligo in 1x T4 DNA ligase buffer (New England BioLabs) was heated to 95°C in a thermocycler for five minutes. The temperature was then reduced by 5°C per minute until a final temperature of 25°C was reached. In parallel, 500 ng



of vector pX330-U6-Chimeric\_BB-CBh-hSpCas9 (pX330; Appendix 7.8.11) [167], which was a gift from Feng Zhang [168], was incubated with 1x CutSmart<sup>TM</sup> buffer (New England BioLabs) and 20 units of restriction enzyme (RE) BbsI-HF<sup>®</sup> (New England BioLabs) at 37°C for 1 hour in a thermoblock. 1  $\mu$ l of the ligated oligo was mixed with 20  $\mu$ l of the linearized pX330 vector as well as 2.5  $\mu$ l 10x T4 ligase buffer and 1.5  $\mu$ l T4 DNA ligase (New England BioLabs). The reaction was incubated at 37°C for 1 hour, and then placed on ice or stored at -20°C.

Transformation of the ligated construct into bacteria was performed according to the following McLab protocol [169], with modifications included: 2  $\mu$ l of the ligated plasmid mixture was added to 50  $\mu$ l of transformation-competent DH5 $\alpha$  *Escherichia coli* (*E. coli*; provided by the Department of Developmental Biology, Georg-August-University of Göttingen). The tube was gently flicked to mix, and incubated on ice for 15 minutes. The cells were then heat-shocked for 45 seconds in a 42°C thermoblock, and subsequently incubated on ice for 2 minutes. 950  $\mu$ l of super optimal broth with catabolite repression (SOC; provided by the Department of Developmental Biology, Georg-August-University of Göttingen) medium was then added to the cells, which were incubated at 37°C for 1 hour in a thermoblock shaking gently at 400 rpm. 100  $\mu$ l of the transformed bacteria was plated using Colirollers plating beads (EMD Millipore Corp.) per 10 cm Lysogeny broth (LB) agar plate (provided by the Department of Developmental Biology, Georg-August University of Göttingen) containing 100  $\mu$ g/ml Ampicillin (Serva), and incubated overnight at 37°C in a bacterial incubator.

The following day, individual colonies were selected with a plastic pipette tip and inoculated into 2 ml LB medium (provided by the Department of Developmental Biology, Georg-August-University of Göttingen) containing 100  $\mu$ g/ml ampicillin, and incubated overnight in a bacterial incubator shaking at 225 rpm. The following day, individual inoculates were centrifuged for 10 minutes at 16,000 xg, RT in a table-top microcentrifuge, and the supernatant was aspirated. Plasmid DNA was purified from the bacterial pellet using a NucleoSpin<sup>®</sup> Plasmid mini-prep kit (Macherey-Nagel) following manufacturer instructions [170], unless the plasmid was to be used immediately for transfections, in which case the NucleoSpin Plasmid Transfection-Grade kit (Macherey-Nagel) was used following manufacturer instructions [171]. The plasmid DNA concentration was determined using a Nanodrop 2000c Spectrophotometer (Thermo Scientific), and the plasmid was sent to Microsynth Seqlab for

Sanger sequencing to validate the gRNA ligation, using the U6 sequencing primer (Appendix 7.10.2.1). Two plasmids containing unique gRNAs were ultimately generated, sequence-validated, and used to target Cas9 to exon 2 of the gene *Tfap2c*, and were designated pX330-a'mTfap2c-exon2-g2 and pX330-a'mTfap2c-exon2-g3.

Blimp1-meGFP (B1) [26] and Stella-eGFP Esg1-tdTomato (SGET) ESCs [172] were transfected as previously described (Section 2.1.3) with one of the two pX330-based gRNA-containing plasmids (pX330-a'mTfap2c-exon2-g2 or pX330-a' mTfap2c-exon2-g3) and plasmid pPY-CAG-mKO-Puro-IP, containing the sequence for the puromycin resistance gene Puromycin N-acetyltransferase derived from *Streptomyces alboniger* (Appendix 7.8.8), and were selected for two days in 1.2  $\mu\text{g}/\text{ml}$  puromycin. Following selection, the pooled transfected cells were single-cell sorted at the fluorescence-activated cell sorting (FACS) SH800S cell sorter (Sony; see Section 2.3) to generate single-colony lines. The knockout lines were characterized by PCR amplification of the region in exon 2 targeted by the gRNA. This reaction was prepared with OneTaq<sup>®</sup> DNA polymerase (New England BioLabs) following manufacturer instructions [173], and included GC enhancer. The following PCR reaction was then run:

1. 94°C for 30 seconds
2. 30 cycles of:
  - 94°C for 30 seconds
  - 56 for 30 seconds
  - 68°C for 30 seconds
3. 68°C for 5 minutes
4. Hold at 8°C

The primers used in this reaction can be found in Appendix 7.10.2.1. The resulting amplicon was cloned into the pMiniT 2.0 vector using the NEB<sup>®</sup> PCR Cloning Kit (New England BioLabs) following manufacturer instructions [174] and cloned into DH5 $\alpha$  transformation-competent bacteria as previously described (Section 2.2.1). Single clones were inoculated, mini-preps were prepared from pelleted inoculates and purified plasmid were Sanger sequenced to verify the presence of a mutation,

as previously described (Section 2.2.1). At least ten plasmid clones were sequenced per KO line to ensure that both alleles were represented. Ultimately, three B1-based AP2 $\gamma$ -KO cell lines were generated in this work: B1 AP2 $\gamma$ -KO gRNA#3 clone #2, B1 AP2 $\gamma$ -KO gRNA#3 clone #3, and B1 AP2 $\gamma$ -KO gRNA#3 clone #11. In this text, they are referred to as lines KO1, KO2, and KO3, respectively. One SGET-based AP2 $\gamma$ -KO cell line was generated in this work: SGET AP2 $\gamma$ -KO gRNA#2 clone #13, named SGET-KO in this text.

## 2.2.2 Homology-directed repair plasmids

In order to report the protein expression of AP2 $\gamma$ , a repair plasmid was designed to allow for the insertion of the *Tdtomato* red fluorescent reporter gene immediately upstream of the *Tfap2c* stop codon, as well as to allow for the insertion of an antibiotic-selection cassette. This repair plasmid was named pmTfap2c-T2A-tdTomato-RoxPGKPuro (pmTfap2c; Appendix 7.8.6). The design of this plasmid was based directly on the plasmid phNANOS3-T2A-tdTomato-RoxPGKPuro (phNANOS3; Appendix 7.8.2) originally published in the Kobayashi *et al.*, 2017 paper [175], and the plasmid was used for cloning with permission from Dr. Toshihiro Kobayashi of the University of Tokyo. The *Tdtomato* and selection cassette sequences of the pmTfap2c plasmid were directly taken from the phNANOS3 plasmid. The generation of this plasmid required the polymerase chain reaction (PCR) amplification of five unique fragments, each starting and ending with 20 nt overlap sequences which were identical to the first 20 nt of the neighboring sequence. The cloning fragments are described in the following table, where size is based on the final length of the fragment post-PCR (and thus including the addition of the Gibson homology arms):

PCR amplification of the fragments was performed with Q5<sup>®</sup> high-fidelity DNA polymerase (New England BioLabs) according to manufacturer protocols [176] using genomic DNA (gDNA) isolated from the B1 ESCs using the DNeasy Blood & Tissue Kit (QIAGEN). Cloning primers required to amplify these fragments are listed in Appendix 7.10.1.2. The prepared PCR reaction was then placed in a thermocycler, and the following PCR program was run:

**Table 1:** DNA fragments required for the construction of plasmid pmTfap2c

<b>Fragment:</b>	<b>Size (bps):</b>	<b>Origin:</b>	<b>Q5 polymerase T<sub>m</sub> (°C):</b>
<i>Tfap2c</i> 5' homology arm	365	Mouse gDNA	61
<i>Tdtomato</i> fluorescent reporter gene	1581	phNANOS3	68
Puromycin/ $\Delta$ TK selection cassette	2566	phNANOS3	60
<i>Tfap2c</i> 3' homology arm	1029	Mouse gDNA	65
MC1-DTA	1266	phNANOS3	61

1. 95°C for 30 seconds
2. 30 cycles of:
  - 98°C for 10 seconds
  - T<sub>m</sub> for 30 seconds
  - 72°C for 30 seconds/kb
3. 72°C for 2 minutes
4. Hold at 8°C

Where T<sub>m</sub> indicates melting temperature, and was determined using the New England BioLabs melting temperature calculator [177]. Once the PCR program was completed, the PCR products were loaded on a 1% agarose gel (1% agarose (Th. Geyer) in 1x Tris-EDTA (TE) buffer; 1 mM EDTA pH 8.0, 10 mM Tris-HCl pH 8.0 [178]), and run for approximately 30 minutes at 100V. The resulting bands were visualized at the LED Blue/White Light Transilluminator (Invitrogen) and cut out with a scalpel. DNA was isolated from the cut gel with the NucleoSpin™ gel and PCR clean-up kit (Macherey-Nagel), and the resulting DNA concentration was determined at the Nanodrop.

In parallel, the plasmid 2c::tdTomato [179] (Appendix 7.8.1) was restriction digested with the enzymes HindIII-HF® (New England BioLabs) and NotI-HF® (New England BioLabs) in 1x CutSmart® buffer for 1 hour at 37°C. Plasmid 2c::tdTomato was a gift from Samuel Pfaff [180]. The restriction digestion reaction was run on a 1% agarose gel for 30 minutes at 100V, and the resulting 5585 bp linearized and digested

band was cut from the gel, and the DNA was purified using the NucleoSpin™ gel and PCR clean-up kit. The linearized backbone and PCR purified fragments were ligated together using a Gibson assembly mastermix (see Appendix 3) [181]. 100 ng of the DNA backbone and each fragment was used for the assembly, with the exception of the *Tfap2c* 5' homology arm fragment, which was less than 1 kb and thus 50 ng was used in the reaction. The reaction proceeded according to the Gibson *et al.*, 2009 published protocol [181]. The ligated plasmid was transformed directly into DH5 $\alpha$  *E. coli* and resulting colonies were inoculated and mini-prepped as previously described (Section 2.2.1). The resulting plasmids were restriction-digested with SnaBI and XhoI in CutSmart® buffer as previously described. When run on a 1% agarose gel at 100V for thirty minutes, plasmids with two appropriately sized bands indicated the correct ligation of the fragments. A plasmid was selected and the plasmid identity was verified with Sanger sequencing. All primers used to validate the correct ligation of the plasmid may be found in Appendix 7.10.2.1.

To allow for the HDR-mediated insertion of the pmTfap2c repair template cassette immediately upstream of the *Tfap2c* stop codon, gRNAs were designed to guide the Cas9 enzyme to the correct genomic location. These gRNAs were designed using the CRISPR guide design tool of the Feng Zhang lab [182], although it should be noted that this tool was deprecated in 2019. These gRNAs were ligated into the pX330 plasmid and sequencing-verified as previously described (Section 2.2.1). Ultimately, one gRNA-containing pX330 plasmid was used for the generation of the *Tfap2c* reporter cell line characterized in this work, and is thus termed pX330-a'mTfap2c-stop-g5. The gRNA oligos used to clone this plasmid may be found in Appendix 7.10.1.1.

B1 ESCs were transfected as previously described (Section 2.1.3) with the repair plasmid pmTfap2c, pPY-CAG-mKO-Puro-IP, and the pX330-based plasmid containing the Cas9 cassette and gRNA targeting the genomic region immediately upstream of the *Tfap2c* stop codon (pX330-a'mTfap2c-stop-g5). The resulting pooled cell lines were termed Blimp1-meGFP AP2 $\gamma$ -tdTomato (BGAT) ESCs, and were selected in 1.2  $\mu$ g/ml puromycin for two days as previously described. The pooled cell lines were single-cell sorted at the SH800S cell sorter (see Section 2.3), and individual clones were assessed for transgenic insertion of *Tdtomato* based on the red fluorescence of the line as observed at the ZEISS Axio Observer.Z1 (Zeiss) microscope. Red fluorescent clones were further transfected with the plasmids pPY-CAG-PBase (Appendix

7.8.9), which encodes the PiggyBac transposase, pPB-CAG-Cas9-IRES-Hygro (Appendix 7.8.7), which encodes the Cas9 enzyme and allows for genomic insertion of the Cas9 gene as well as a *E. coli* aminoglycoside phosphotransferase hygromycin resistance gene (Hygro) which is separated from the Cas9 by an internal ribosome entry site (IRES), and pCAGGS-DRE-IH (Appendix 7.8.3), which encodes the Dre recombinase gene from a P1-related phase [183] which allows for the excision of the Rox-flanked selection cassette downstream of *Tdtomato*, and selected the cells in 200  $\mu\text{g}/\text{ml}$  Hygromycin for five days. The pooled lines were then single-cell sorted at the SH800S cell sorter, and individual lines were genotyped via PCR and reporter fluorescence was characterized via indirect immunofluorescent (IF) staining and imaging. Primers used to genotype the BGAT line may be found in Appendix 7.10.2.1.

## 2.3 FACS

To prepare for fluorescence activated cell sorting (FACS), ESCs and EpiLCs were dissociated as previously described (Section 2.1.1). After dissociation and resuspension in passaging medium, the cell suspension was centrifuged for 4 minutes at 1,200 xg, RT. The pellet was then gently resuspended in FACS buffer (2% FBS in PBS) containing 1:1000 4',6-diamidino-2-phenylindole (DAPI; Roche Diagnostics GmbH), a stain which fluoresces when bound to adenine - thymine (A-T) interactions in double-stranded DNA [184]. Strong DAPI fluorescence in sorted cells indicates cell death, as dead or apoptotic cells have compromised cellular membranes and thus allow DAPI to freely stain double-stranded DNA within the nuclei. Thus, DAPI was included in the FACS buffer to indicate cell death and allow for the exclusion of dead cells from sorted populations. After resuspension of the cell pellet in FACS buffer, the cell suspension was passed through a 35  $\mu\text{m}$  nylon-mesh cell strainer (Falcon) and placed on ice until it could be sorted at the SH800S cell sorter (Sony).

Dissociation of EBs for FACS analysis depended upon the method of EB generation. If EBs were grown in 96-well microtiter plates according to the Hayashi et al. [21] protocol, all EBs of a given condition were collected in one 1.5 ml microreaction tube and briefly spun in a benchtop microcentrifuge to collect EBs at the bottom of the tube. The medium was then aspirated and the EBs were washed with 500  $\mu\text{l}$  PBS. If EBs were grown according to the Gruhn and Günesdogan [162] protocol, the contents of each microwell were pipetted into a 15 ml tube. The empty well was gently washed with 2 ml -cyto GK15 to ensure all EBs had been dislodged from the microwells, and

the medium was pipetted into the same 15 ml tube, bringing the total volume to 4 ml. The 15 ml tubes were not disturbed for 5 minutes at RT to allow the EBs to collect at the bottom of the tube. The supernatant was then aspirated, and the EBs were washed with 1 ml 1x PBS and transferred to a 1.5 ml microreaction tube. From this point on, the protocol for dissociation of EBs is identical regardless of EB generation protocol. The tubes were briefly spun on a benchtop microcentrifuge, and the supernatant was aspirated. The EBs were resuspended in 250  $\mu$ l TrypLE<sup>TM</sup> Express and incubated in a 37°C heat block shaking at 400 rpm. Day 2 (d2) EBs incubated for 4 minutes, d4 EBs incubated for 6 minutes, and d6 EBs incubated for 8 minutes in total. Halfway through the incubation of the d4 and d6 EBs at the heat block, the microtubes were gently flicked to dislodge the pellet at the bottom of the tube. After the incubation was complete, 500  $\mu$ l passaging medium was added to each tube, and the cell suspension was gently pipetted up and down until large aggregates of cells had dissociated. The suspension was centrifuged for 4 minutes at 1,200 xg, RT. The supernatant was aspirated, and cells were resuspended in FACS buffer containing 1:1000 DAPI and passed through a cell strainer. The cells were placed on ice until sorting could proceed.

At the FACS, cells were gated first through side-scatter (SSC) against forward scatter (FSC) detectors, which allows one to discriminate cells based upon size. The cells were next gated through FSC width against SSC height detectors, which allows one to exclude doublets from sorting. The cells were next gated through FSC against DAPI detectors, which allows for the exclusion of highly DAPI+ cells, indicating cell death. The resulting cell population was used for analysis and sorting, and was gated against GFP and/or tdTomato detectors to identify the representative fluorescent cell populations. When sorting cells based on fluorescence, E14 murine non-fluorescent ESCs [185] were always included as a control.

Cell sorting was used for two separate functions: 1) Pooled, transfected cells were sorted in single-cell sort mode such that one cell was sorted per well of a Nunc<sup>TM</sup> MicroWell<sup>TM</sup> 96-Well, nunclon delta-treated, flat-bottom microplate (Thermo Scientific), pre-coated with 0.1% gelatin (Merck) for 10 minutes at RT, and filled with 100  $\mu$ l 2i+LIF medium. 2) Cells were two-way sorted based upon fluorescence, such that two discrete populations were identified and sorted into 1.5 ml microreaction tubes for analysis or further experimentation. If cells were taken for RNA isolation, the microtubes were immediately flash frozen in liquid nitrogen and stored at -80°C. If

cells were taken for ATAC-seq (Section 2.8), they were placed on ice and immediately processed. If cells were taken during the *Tfap2c* tiling screen (Section 2.9.4), the cell suspension was centrifuged for 4 minutes at 1,200 xg, RT, and the supernatant aspirated. The pellet was then frozen at -20°C until the gDNA was to be isolated. To determine significance between sorted cell populations during analysis, unpaired t-tests were performed using Prism 9 software (GraphPad).

## 2.4 qRT-PCR

Quantitative reverse-transcription polymerase chain reaction (qRT-PCR) was performed to measure the expression of specific messenger ribonucleic acid (mRNA) transcripts in a given condition. First, RNA was isolated from a given sample using the RNeasy<sup>®</sup> plus micro kit (QIAGEN). The concentration of RNA was measured using a NanoDrop<sup>™</sup> spectrophotometer. 1  $\mu$ g of ESC or EpiLC RNA or 40 ng RNA isolated from EBs was reverse transcribed into complementary DNA (cDNA) using a Quantitect<sup>®</sup> Reverse Transcription kit (QIAGEN) following manufacturer instructions [186]. A 1:10-1:20 dilution of cDNA was used to prepare a qRT-PCR reaction (4  $\mu$ l 2x KAPA SYBR<sup>®</sup> FAST qPCR Master Mix (Kapa Biosystems), 0.08  $\mu$ l 10  $\mu$ M forward primer, 0.08 10  $\mu$ M  $\mu$ l reverse primer, 1  $\mu$ l diluted cDNA, 2.84  $\mu$ l dH<sub>2</sub>O), and the prepared reaction was pipetted into the wells of a 96-well clear Multiplate<sup>®</sup> PCR plate<sup>™</sup> (Bio-Rad Laboratories) and sealed with a Microseal<sup>®</sup> 'B' seal (Bio-Rad Laboratories). All qPCR primers used in this work were validated for efficiency, and may be found in Appendix 7.10.2.3.

The qPCR reaction was run at the CFX96<sup>®</sup> Real-Time PCR detection system (Bio-Rad Laboratories) with the following cycle:



1. 95°C for 3 minutes
2. 40 cycles of:
  - 95°C for 3 seconds
  - 60°C for 25 seconds
3. Melt curve for 25 minutes:
  - 65°C for 10 seconds
  - Temperature raised by 0.2°C/10 seconds
  - Final temperature 95°C
4. Hold at 4°C

The resultant cycle threshold (Ct) values generated during the qRT-PCR run were used to determine the  $\Delta Ct$  values, indicating expression of a target mRNA relative to the control gene used in this study, *Arbp* (Attachment region binding protein gene).  $\Delta\Delta Ct$  values were calculated to indicate expression of a target mRNA in an experimental sample relative to a control sample. Fold change was calculated to determine the fold difference in expression of  $\Delta\Delta Ct$  in an experimental sample relative to a control sample. Standard deviation was calculated for each sample and included in the plotted data. The methods for calculating these values are written below, and are taken from a Smith College guide to qRT-PCR analysis [187]:

$$\Delta Ct = \frac{\text{Avg. target gene Ct}}{\text{Avg. housekeeping gene Ct}} \quad \Delta\Delta Ct = \frac{\Delta Ct \text{ target gene}}{\Delta Ct \text{ housekeeping gene}}$$

$$\text{Fold change} = 2^{-(\Delta\Delta Ct)}$$

$$\text{Variance} = ((\text{StDev. Housekeeping Gene})^2(\text{StDev. Target Gene})^2)^{1/2}$$

Where StDev is standard deviation as calculated using Microsoft Excel (Microsoft). Significance between samples was determined based on fold-change as well as standard deviation, and was identified by first determining the variance of the samples via the t-test function, then determining the significance of the samples using the generated variances in the z-test function of Microsoft Excel.

## 2.5 Indirect immunofluorescence

### 2.5.1 IF staining of cell monolayers

Indirect immunofluorescence (IF) staining was employed to image cellular proteins in fixed samples of adherent cells, and was performed following a modified unpublished protocol developed by Dr. Petra Hajkova of Imperial College London [188]. In order to perform ICC, ESCs or EpiLCs were plated in their respective media in 4-well Nunc™ plates (Thermo Scientific) on glass coverslips (Hecht Karl) coated with 1:60 fibronectin in PBS for at least one hour at 37°C. When the cells were ready for imaging, they were washed 1x in PBS and subsequently fixed for 10 minutes in 4% paraformaldehyde (Agar Scientific) in PBS. The slides were subsequently washed 3x in PBS and incubated for thirty minutes at RT in permeabilization buffer [188], containing 0.1% Triton® X-100 (Fisher Bioreagents) and 1% Albumin fraction V (Carl Roth). The slides were then incubated overnight at 4°C in primary antibody (Appendix 7.2) diluted in permeabilization buffer. The following day, the slides were washed 3x with PBS and incubated for one hour at RT in the dark in secondary antibody (Appendix 7.2) with 1  $\mu$ g/ml DAPI diluted in permeabilization buffer, after which they were washed 3x in PBS and mounted with VECTASHIELD® PLUS antifade mounting medium (Vector Laboratories, Inc.) on 90° ground frosted microscope slides (Menzel Gläser) and sealed with clear nail polish. The slides were dried for 10 minutes at RT in the dark, and then stored at 4°C until imaging.

### 2.5.2 IF staining of cryosections

IF staining was employed to image cellular proteins of fixed, cryosectioned EBs, and was performed using a published protocol [189] with modifications as follows: First, EBs were collected in a well of a 96-well flat-bottom plate and the medium was gently removed. The EBs were washed 1x with PBS, and then fixed in 4% PFA for 15 minutes at RT. The EBs were then washed 3x with PBS, and subsequently incubated in 10% D(+)-Saccharose (Carl Roth) in PBS for thirty minutes at RT. The solution was then gently aspirated, and the EBs were further incubated in 20% saccharose in PBS for thirty minutes before a final thirty minute incubation in 30% saccharose in PBS. The EBs were then gently transferred to a 7 x 7 disposable plastic base mold (EpreDia). The supernatant was gently removed, and the EBs were submerged in OCT embedding matrix (Carl Roth) and gently rocked for 15 minutes at RT. The molds were then placed on dry ice until frozen, and stored at -80°C until further use.

The embedded EBs were cryosectioned at a Leica CM1950 microtome (Leica) into sections of eight microns and placed on Superfrost<sup>TM</sup> plus adhesion microscope slides (EpreDia). The slides were then stained according to the Hajkova protocol [188]: A lipid barrier pen (Vector Laboratories) was used to draw an impermeable barrier around the sections on the slide, and they were then washed in 3x for five minutes in PBS in a Coplin jar. The slides were placed in a humidified staining chamber and incubated with permeabilization buffer for thirty minutes at RT. The buffer was gently removed, and the slides were stained overnight in the humidified chamber at 4°C in primary antibody (Appendix 7.2) diluted in permeabilization buffer, with 24 x 40 mm coverslips (Menzel Gläser) gently placed on top of the slides to prevent dehydration. The following day, the slides were washed 3x five minutes in PBS in Coplin jars. The slides were then returned to the humidified chamber and stained with secondary antibody (Appendix 7.2) and 1  $\mu\text{g}/\text{ml}$  DAPI in permeabilization buffer for 1 hour. The slides were then washed 1x for five minutes at RT in the dark humidified chamber with permeabilization buffer, before being mounted with VECTASHIELD<sup>®</sup> PLUS, covered with a 24 x 40mm glass coverslip, and sealed with clear nail polish. The slides were allowed to dry for ten minutes at RT in the dark, and afterwards were stored at 4°C until imaging.

### 2.5.3 Image analysis

Light microscopy images of live cells were taken with a ZEISS Axio Observer.Z1 microscope. Confocal images were taken at the ZEISS LSM 980 (ZEISS) confocal microscope. The resulting images were uniformly adjusted and analyzed using ZEISS Zen software (ZEISS) and Affinity Designer (Serif (Europe), Ltd.).

## 2.6 Western blotting

ESC and EpiLC cell lysates were prepared according to an unpublished protocol developed by Dr. Lena Wartosch at the Max Planck Institute for Multidisciplinary Sciences [190]: First, adherent cells grown in 6-well Nunc<sup>TM</sup> cell-culture treated multidishes (Thermo Scientific) or 10 cm Nunc<sup>TM</sup> EasYDish<sup>TM</sup> dishes (Thermo Scientific) cell culture plates were washed with 1x PBS. After the PBS was removed, either 75  $\mu\text{l}$  (per well of 6-well plate) or 200  $\mu\text{l}$  (per 10 cm plate) lysis buffer (50 mM Tris-HCl pH 7.4, 50 mM NaCl (AppliChem), 2% sodium dodecyl sulfate (SDS; Carl Roth), and 1 tablet cComplete<sup>TM</sup> mini EDTA-free protease-inhibitor tablet protease inhibitor (Roche) in  $\text{dH}_2\text{O}$ ) was added directly to the cells. The cells were scraped from the

surface of the plate using a Cell Scraper S (TPP), and pipetted into a QIAshredder column (QIAGEN). The lysate was centrifuged once for two minutes at 16,000 xg, RT, and then reloaded directly onto the same column and centrifuged again following the same specifications. The lysate could then be directly flash-frozen in liquid nitrogen and stored at -80°C for later use, or placed on ice for immediate use.

Protein concentration was determined using the Qubit™ Protein and Protein Broad Range (BR) Assay Kit (Invitrogen) with the Qubit® 2.0 fluorometer (Invitrogen) system following manufacturer protocols [191]. Once the concentration had been determined, an appropriate quantity of protein was mixed with 5  $\mu$ l Laemmli sample buffer (Bio-Rad Laboratories) and dH<sub>2</sub>O *ad* 20  $\mu$ l and boiled at 95°C for 5 minutes. The samples could then be stored at -20°C until use.

Prepared protein lysates were pipetted into the wells of a 4% SDS polyacrylamide (SDS-PAGE) stacking gel and subsequently 10% sodium dodecyl-sulfate polyacrylamide gels (Appendix 7.6.1) at 100V in a Mini Trans-Blot® cell (Bio-Rad Laboratories) until the dye front of the Laemmli buffer reached the bottom of the gel. 5 $\mu$ l of PageRuler™ Prestained Protein Ladder (Thermo Scientific) was always run in an empty lane of the gel as a reference for protein size in kilodaltons (kDa). The transfer cassette was then assembled, comprising two filter pads, two filter papers, and an Amersham™ Protran™ 0.45  $\mu$ m NC Nitrocellulose Blotting Membrane (Cytiva) pre-soaked in 1x transfer buffer (192 mM glycine (Carl Roth), 25 mM Tris-HCL pH 7.6, 20% methanol (VWR) [192]) carefully and directionally layered with the gel to ensure the correct transfer of proteins. The transfer ran for 1 hour at 4°C, 100v with an ice pack, following manufacturer instructions [193]. The blot was then removed and the excess membrane trimmed with a scalpel. The blot was incubated in 5% milk buffer (5% skim milk (BD) and 0.1% Tween-20 (Fisher Bioreagents) in 1x PBS) for 30 minutes at RT, gently rocking. The blot was then incubated overnight with primary antibody (Appendix 7.2) in 5% milk buffer at 4°C in a 50 ml tube, continually rolling. The next day, the blot was washed 3x for five minutes with PBS-T (0.1% Tween-20 in 1x PBS). The blot was subsequently incubated with horseradish-peroxidase-conjugated (HRP) secondary antibody (Appendix 7.2) in 5% milk buffer for one hour at RT in a 50 ml tube, continually rolling. The blot was washed 3x five minutes with PBS-T, and then incubated with Pierce™ ECL Western Blotting Substrate (Thermo Scientific) for approximately two minutes at RT, before being imaged for chemiluminescence at the ChemiDoc™ Imaging System (Bio-Rad Laboratories).

Once the blot had been imaged for chemiluminescence, a colorimetric image was taken and merged using the ChemiDoc<sup>TM</sup> Image Lab Touch software (Bio-Rad) or Affinity Designer.

## 2.7 RNA-seq

### 2.7.1 RNA-seq library preparation and sequencing

RNA was isolated for RNA-sequencing (RNA-seq) library preparation using the RNeasy<sup>TM</sup> RNA isolation kit. Before RNA-seq library prep was performed, the quality of RNA was assessed using a Fragment Analyzer (Agilent), which generates a RNA Quality Number (RQN) as an indication of RNA quality, with values ranging from 1 (completely degraded) to 10 (ideal) [194]. Any RNA sample with an RQN < 7.0 was not used for RNA-seq library preparation. RNA was collected from ESC, EpiLC, and d4/d6 GFP+ FACS-sorted samples. 3x WT samples were collected for each cell state condition, while one sample was collected for each AP2 $\gamma$  KO line, however three separate lines were used per condition. Thus, the RNA input used to generate the RNA-seq libraries comprised 3x WT samples and 3x separate KO lines per condition.

The RNA sample concentration was measured using a Qubit<sup>®</sup> 2.0 fluorometer and Qubit<sup>®</sup> high-sensitivity RNA assay kit (Life Technologies) following manufacturer instructions [195]. 100 nanograms (ng) of RNA per sample was used for input with the NEBNext<sup>®</sup> Ultra<sup>TM</sup> II Directional RNA library prep kit for Illumina<sup>®</sup> (New England BioLabs). The protocol provided by the manufacturer [196] was followed exactly, with the sole user specifications being that 14 cycles were chosen for PCR library amplification following adaptor ligation in 1:25 diluted adaptor. Single index Illumina<sup>®</sup> primers were selected from NEBNext<sup>®</sup> Multiplex Oligos for Illumina<sup>®</sup> primer sets 2-3 (New England BioLabs). All sequencing primers may be found in Appendix 7.10.2.2. The libraries were single left-sided bead purified an additional time with 1x Agencourt<sup>®</sup> RNAClean<sup>®</sup> XP Beads (Beckman Coulter) following manufacturer instructions [197], and then re-analyzed at the Fragment Analyzer to ensure that the libraries had correctly amplified and no primer or adapter contamination remained.

Two of the libraries (samples B1B\_ESC and B1D\_ESC, renamed for clarity and sim-

plicity in this work ESC WT1 and ESC WT3) were sequenced on a NovaSeq 6000 (Illumina) for paired-end 100 bp reads, however the rest of the RNA-seq libraries were sequenced on a NovaSeq 6000 for single-end 100 bp reads. These libraries, originally named B1C\_ESC, B3.2\_ESC, B3.3\_ESC, B1A\_EpiLC, B1B\_EpiLC, B1C\_EpiLC, B3.2\_EpiLC, B3.3\_EpiLC, B3.11\_EpiLC, B1A\_d4, B1B\_d4, B1C\_d4, B3.2\_d4, B3.3\_d4, B3.11\_d4, B1A\_d6, B1B\_d6, B1C\_d6, B3.2\_d6, B3.3\_d6, B3.11\_d6 were renamed to WT2 ESC, KO1 ESC, KO2 ESC, KO3 ESC, WT1 EpiLC, WT2 EpiLC, WT3 EpiLC, KO1 EpiLC, KO2 EpiLC, KO3 EpiLC, WT1 d4 GFP+, WT2 d4 GFP+, WT3 d4 GFP+, KO1 d4 GFP+, KO2 d4 GFP+, KO3, d4 GFP+, WT1 d6 GFP+, WT2 d6 GFP+, WT3 d6 GFP+, KO1 d6 GFP+, KO2 d6 GFP+, and KO3 d6 GFP+, respectively, for clarity and simplicity. Total reads from the single-end libraries ranged from approximately 36 million (Mio) to 58 Mio reads, while paired-end reads comprised 46 Mio fragments (ESC WT1) and 38 Mio fragments (ESC WT3). Only one pair of the paired-end reads was used for sequencing analysis, and the files were thus treated as single-end.

## 2.7.2 RNA-seq data analysis

RNA-seq data was largely analyzed using the Galaxy Browser [198]. First, RNA-seq read quality was assessed using the tools FastQC [199] and the quality report aggregator MultiQC [200]. 3' adapter sequences were trimmed from the reads using CutAdapt [201] with the reference Illumina adapter sequence of AGATCGGAA-GAGCACACGTCTGAACTCCAGTCAC, and reads with a quality score below 20 and a total length below 20 were filtered from the trimmed reads. and the quality of the trimmed reads was re-assessed using FastQC. The trimmed reads were then aligned to the mm10 genome using HISAT2 [202] and duplicates were marked with the Picard [203] tool MarkDuplicates. Count tables were generated from the aligned reads using featureCounts [204], with the reverse stranded specification included. The subsequent count tables were annotated with annotateMyIDs [205]. Differential expression between samples was quantified using DESeq2 [206]. Differentially expressed genes were visualized using heatmap2 [207], by which the data was clustered by column and scaled by row. Gene ontology was performed on differentially expressed gene sets with Goseq [208]. pcaExplorer [209] was used to generate principal component analysis (PCA) plots and Euclidian hierarchical clustering analysis heatmaps to better assess the quality of RNA-seq data at the counts level, as well as to produce plots of normalized read counts of genes.

As a quality control step, the Read Distribution [210], Gene Body Coverage [210], IdxStats [211], and Infer Experiment [210] tools were used to examine: the read distribution across features, the distribution of reads along the gene body from 5' to 3' as an indication of degradation, the number of reads mapped per chromosome, and the strandedness of the libraries, respectively. RNA-seq BAM files were converted to Bigwig files for visualization with the Interactive Genomics Viewer (IGV; [212]) using the deepTools bamCoverage tool [213].

## 2.8 ATAC-seq

Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq) was performed to identify changes in chromatin accessibility between AP2 $\gamma$ -KO and WT cells, as well as to compare the WT cell states ESCs, EpiLCs, and PGCLCs. To this end, 3x B1 WT samples were used per cell state condition, while 2x AP2 $\gamma$ -KO samples were used, each deriving from an individual clonal AP2 $\gamma$ -KO line (lines KO1-2). The ATAC-seq protocol was modified by Dominik Mühlen of the University of Göttingen from the original 2015 Buenrostro *et al.* protocol [214], in short: 5 x 10<sup>4</sup> ESCs or EpiLCs, or 2.5 x 10<sup>4</sup> d4 or d6 GFP<sup>+</sup> FACS-sorted PGCLCs were washed 1x with 50  $\mu$ l ice-cold PBS, and centrifuged for five minutes at 500 xg, 4°C. The cell pellet was gently resuspended in ATAC lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub> (Carl Roth), and 0.1% IGEPAL CA-630 (Sigma Aldrich) in dH<sub>2</sub>O). The cell suspension was immediately centrifuged for ten minutes at 500 xg, 4°C. The supernatant was discarded, and the pellet was resuspended in transposition reaction: 25  $\mu$ l Nextera Transposition buffer (Illumina), 2.5  $\mu$ l Nextera Tn5 Transposase (Illumina), and 22.5  $\mu$ l dH<sub>2</sub>O for the 5 x 10<sup>4</sup> ESCs or EpiLCs, or 12.5  $\mu$ l Nextera Transposition buffer (Illumina), 1.25  $\mu$ l Nextera Tn5 Transposase (Illumina), and 11.25  $\mu$ l dH<sub>2</sub>O for the 2.5 x 10<sup>4</sup> d4/d6 GFP<sup>+</sup> cells. The reactions were incubated at 37°C for 75 minutes in a thermocycler.

The transposed library fragments were PCR amplified in a reaction containing the transposed DNA, 25  $\mu$ M forward custom Nextera primer, 25  $\mu$ M reverse custom Nextera primer, 1x NEBNext<sup>®</sup> High Fidelity PCR master mix (New England Biolabs), and dH<sub>2</sub>O *ad* 50  $\mu$ l. All custom Nextera sequencing primers may be found in Appendix 7.10.2.2. The PCR reaction was performed according to the following cycle:

1. 98°C for 30 seconds
2. 12 cycles of:
  - 98°C for 10 seconds
  - 63°C for 30 seconds
  - 72°C for 1 minute
3. Hold at 4°C

To determine the correct number of cycles for amplification, the first time this PCR was run, after 5x cycles the PCR was aborted and placed on ice. 5  $\mu$ l of the reaction was taken for quantitative PCR (qPCR) amplification to examine the amplification curve in real-time and calculate the ideal total number of amplification cycles for the library PCR. The qPCR reaction was composed of 5  $\mu$ l PCR reaction template, 0.25  $\mu$ l forward custom Nextera primer, 0.25  $\mu$ l reverse custom Nextera primer, 0.09  $\mu$ l SYBR<sup>®</sup> Green 1 nucleic acid stain (Lonza), 5  $\mu$ l NEBNext<sup>®</sup> High-Fidelity 2x PCR master mix, and 4.41  $\mu$ l dH<sub>2</sub>O. The qPCR reaction was run at the CFX96 Real-Time cycler with the following parameters:

1. 98°C for 30 seconds
2. 10 cycles of:
  - 98°C for 10 seconds
  - 63°C for 30 seconds
  - 72°C for 1 minute

The reaction cycle threshold (Ct) values were plotted against the cycle number to produce a graph. The cycle number which was equivalent to 1/3 the total maximum fluorescent intensity was then selected as the total number of cycles necessary for library preparation, in this case twelve total cycles for ATAC-seq library PCR. Once all libraries had been amplified, they were purified using the MinElute<sup>®</sup> PCR Purification Kit (QIAGEN) and eluted in 20  $\mu$ l elution buffer. The libraries were 1.8x single left-sided bead purified following manufacturer instructions [197] and the libraries were visualized using a TapeStation D1000 fragment analysis device (Agilent) to examine the libraries for expected size distribution and presence



of nucleosomal ladder peaks. The libraries were subsequently underwent paired-end 50 bp sequencing on a NovaSeq 6000 sequencer. The resulting sequenced libraries resulted in between 23 Mio and 85 Mio fragments, per sample. The sequencing samples were originally named B1A\_ESC, B1B\_ESC, B1C\_ESC, B3.2\_ESC, B3.3\_ESC, B1A\_EpiLC, B1B\_EpiLC, B1C\_EpiLC, B3.2\_EpiLC, B3.3\_EpiLC, B1A\_d4, B1B\_d4, B1C\_d4, B3.2\_d4, B3.3\_d4, B1A\_d6, B1B\_d6, B1C\_d6, B3.2\_d6, B3.3\_d6, and were renamed to WT1 ESC, WT2 ESC, WT3 ESC, KO1 ESC, KO2 ESC, WT1 EpiLC, WT2 EpiLC, WT3 EpiLC, KO1 EpiLC, KO2 EpiLC, WT1 d4 GFP+, WT2 d4 GFP+, WT3 d4 GFP+, KO1 d4 GFP+, KO2 d4 GFP+, WT1 d6 GFP+, WT2 d6 GFP+, WT3 d6 GFP+, KO1 d6 GFP+, and KO2 d6 GFP+, respectively, for clarity and simplicity.

Using the Scientific Computing Cluster (SCC) hosted by the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) for the University of Göttingen, the quality of the reads of the ATAC-seq libraries was assessed using FastQC and the quality score aggregator MultiQC. Nextera adapter sequences were trimmed from the reads using the tool Trim Galore! [215], and the quality of the trimmed reads was again assessed using FastQC and MultiQC. The trimmed reads were aligned to the mm10 genome using Bowtie2 [216] with the specifications: `--very-sensitive`, up to 10 alignments per read (`-k 10`), and the maximum fragment length for reads was set to 2000 bps (`-X 2000`). The alignment success was examined using FastQC, and mitochondrial reads were removed from the libraries. Duplicates were marked and removed using the Picard MarkDuplicates tool, and SAMtools [211] was used to filter out reads with a quality score below 30 (`-q 30`), reads with the flag 1024 indicating unmapped, optical duplicate, or incorrectly paired reads (`-F 1024`), and reads with the flag 1804 indicating unmapped, improperly aligned reads which fail quality checks (`-F 1804`). The deepTools [213] tool AlignmentSieve was used to shift reads to accurately center the transposon site. The BAM files containing the shifted reads were converted to BigWig files for visualization on IGV using the deepTools bamCoverage tool. The shifted reads were used to call peaks with Macs2 [217] with the callpeak function and the specifications: BED file as input file (`-f BED`), mm10 as reference genome (`-g mm`), no Macs2 model should be built (`--nomodel`) and duplicates are kept (`--keep-dup all`).

To find differentially accessible peaks (DAPs) between cell types and between KO and WT states, the peak files of all samples of the same cell type and KO status,

*i.e.* all WT ESC peak files, were merged using a custom R (R Core Team) [218] script designed by Xiaojuan Li of Georg-August University of Göttingen. This script identified the coordinates of all peaks which were present in 2/3 or 2/2 input files, and output a file comprised of the identified coordinates. The WT and KO merged peak files for each given cell state were concatenated to generate a master merged peaks file of all peaks which were present in 2/3 WT samples and 2/2 KO samples. This file was used as the reference file with the featureCounts tool to count which coordinates identified in the merged peaks files were present in the shifted reads BAM files. The resulting count tables were used to identify DAPs between WT and KO samples within a particular cell type using the EdgeR [219; 220] tool on the Galaxy browser, with the specifications: use robust settings, adjusted p-value of  $< 0.05$  for significance, and use the likelihood ratio test. The resulting DAPs were genomically annotated using the hypergeometric optimization of motif enrichment (HOMER) [221] tool annotatePeaks.pl, using the specifications: mm10 as reference genome and perform gene ontology analysis (-go). The HOMER tool findMotifsGenome.pl was used to identify TF motif enrichment in DP files [221], with the specifications to use mm10 as reference and to use the size of the ATAC fragments directly via “size given”. The HOMER tool getDifferentialPeaks was used to generate overlapping peak data [221], which was visualized for quality control purposes with UpSetR [222] using a script developed by Stephen Kelly [223].

## 2.9 *Tfap2c* tiling screen

### 2.9.1 *Tfap2c*-screen gRNA library design

The gRNA sequences used in the *Tfap2c* tiling screen gRNA library were primarily generated by Xiaojuan Li at Georg-August University of Göttingen using the count\_spacers.py script developed by Joung *et al.* [156; 224] with the following modifications: the maximum number of guides selected per gene (-n) was set to zero and the minimum spacing required between cleavage sites (-s) was set to one nt. All resultant gRNAs with an off-target score  $> 0.50$  were selected. gRNAs targeting the negative control ESC enhancers, 5 kb negative control genomic regions, and *Tdtomato* were generated using the Benchling CRISPR guide RNA design tool [225]. The six *Tdtomato*-targeting gRNAs with the highest on- and off-target scores were chosen for use, while the top one hundred scoring gRNAs were chosen per each targeted genomic region. Individual gRNAs targeting control mouse genes were taken from the GeCKO mouse library, version 2 [153], and are named

in the original library as follows: *Tfap2c* (MGLibA\_53528-30, MGLibB\_53512-14), *Prdm1* (MGLibA\_42592-4, MGLibB\_42579-81), *Prdm14* (MGLibA\_42583-5, MGLibB\_42571-3), *Dppa3* (MGLibA\_14830-32, MGLibB\_14823-5), and *Dppa5a* (MGLibA\_14836-8, MGLibB\_14829-31). One thousand scrambled gRNAs were also taken from the GeCKO mouse library, version 2, and are identified as MGLibA\_66406 - 67405 in the GeCKO library. Six *Egfp*-targeting gRNAs were taken from the 2014 Shalem *et al.* publication [226]. The final library comprised 30,714 unique gRNAs. All information concerning regions used for gRNA design, origin of gRNAs, responsible gRNA designer, and purpose of gRNA may be seen in the table in Appendix 7.11. The gRNA library was ordered from Genscript with the forward adapter sequence:

5' - GCAGATGGCTCTTTGTCCTAGACATCGAAGACAACACCGN<sub>20</sub>GTTTTAGTCTTCTCGTCGC - 3'

where N<sub>20</sub> represents the unique twenty nucleotide gRNA sequence. This adapter sequence allows for the amplification and insertion of the BbsI-HF-processed gRNA sequence into the CRISPR gRNA expression plasmid pKLV2-U6gRNA5(BbsI)-PGK puro2ABFP-W (Appendix 7.8.4) as described by Tzelepis *et al.*, 2016 [154]. Plasmid pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W (pKLV2-W) was a gift from Kosuke Yusa [227].

### 2.9.2 *Tfap2c*-screen gRNA library generation

Preparation of the gRNA library for insertion into the pKLV2-W plasmid followed the protocol from Tzelepis *et al.*, 2016 [154], and is described below with modifications included:

100  $\mu$ l of NEBNext<sup>®</sup> Ultra<sup>™</sup> II Q5<sup>®</sup> Master Mix (New England BioLabs) was mixed on ice with 8  $\mu$ l of the gRNA oligo pool (final concentration of 0.04 ng/ $\mu$ l), 2.5  $\mu$ l primer 79-mer U1 (final concentration of 0.5  $\mu$ M), 2.5  $\mu$ l primer 79-mer L1 (final concentration of 0.5  $\mu$ M), and dH<sub>2</sub>O *ad* 200  $\mu$ l. Primers used during the generation of this library maybe found in Appendix 7.10.2.2. The mixture was pipetted into 25  $\mu$ l aliquots in PCR strips and placed in the thermocycler with the following protocol:

1. 98°C for 30 seconds
2. 20 cycles of:
  - 98°C for 10 seconds
  - 63°C for 10 seconds
  - 72°C for 15 seconds
3. 72°C for 2 minutes
4. Hold at 4°C

The PCR reactions were then pooled and purified with the QIAquick Nucleotide Removal Kit (QIAGEN) following manufacturer instructions [228]. The resulting PCR product was run in a 10% Tris-Borate-EDTA (TBE) gel (Appendix 7.6.2) in a Mini-PROTEAN<sup>®</sup> Tetra gel electrophoresis chamber at 100V for approximately one hour. The gel was soaked in 1:10,000 SYBR<sup>™</sup> Safe DNA Gel Stain (Invitrogen) for 10 minutes at RT in the dark, and was subsequently visualized on the blue light transilluminator. The digested bands were cut from the gel with a scalpel, and the DNA was removed from the gel using the Lonza Group crush and soak procedure [229], in brief: the TBE gel containing the DNA was cut and crushed using a scalpel. One to two volumes of TE buffer was added and the gel was incubated for four hours at 37°C, and subsequently centrifuged for one minute at 4°C at 12,000 xg in a tabletop centrifuge. The supernatant was gently removed and placed in a fresh tube, and 0.5 volumes of TE buffer was added to the crushed gel and again centrifuged. The supernatant was removed and added to the previously collected supernatant, and the gel was discarded. Gel fragments were removed by running the supernatant through a NucleoSpin<sup>™</sup> gel and PCR clean-up kit column, and two volumes of cold ethanol was then added to the flow-through. The solution was incubated on ice for thirty minutes, and subsequently centrifuged for ten minutes at 4°C, 12,000 xg. The supernatant was carefully removed, and the pellet was resuspended in 200  $\mu$ l TE buffer. 25  $\mu$ l of 3M sodium acetate (Alfa Aesar), pH 5.2 was added to the solution, after which two volumes of cold ethanol was again added, and the incubation and subsequent centrifugation proceeded as previously described. The resultant pellet was washed once in 70% ethanol before air drying for ten minutes, and was subsequently dissolved in 10  $\mu$ l TE buffer.

In parallel, 25  $\mu$ l CutSmart<sup>™</sup> 10x buffer was mixed with 5  $\mu$ l BbsI-HF<sup>®</sup>, 5  $\mu$ g pKLV2-

W plasmid, and dH<sub>2</sub>O *ad* 250  $\mu$ l. Once the reaction had been thoroughly mixed, it was aliquoted into 50  $\mu$ l reactions and incubated at 37°C in a thermoblock for 1 hour. A control restriction-digestion proceeded in parallel following the same specification however using BbsI-HF<sup>®</sup> and BamHI-HF<sup>®</sup> restriction digestion enzymes in order to generate a replication-incompetent ligation control. The restriction-digested plasmids were then run on a 1.5% agarose gel at 100V for 30 minutes, and the resulting gel was visualized on the blue light transilluminator to identify the 8.6 kb band of the linearized pKLV2-W plasmid. The band was cut from the gel and the DNA was purified using the NucleoSpin<sup>™</sup> gel and PCR clean-up kit following manufacturer instructions [230].

Ligation of the gRNA library insert into the linearized pKLV2-W plasmid occurred according to the T4 DNA ligase manufacturer's protocol [231], and was prepared as follows: 200 ng of the RE-digested, purified gRNA library and 1.32  $\mu$ g of purified, digested pKLV2 plasmid were mixed with 32  $\mu$ l 10X T4 ligase buffer, 18  $\mu$ l T4 DNA ligase, and dH<sub>2</sub>O *ad* 360  $\mu$ l. The reaction was incubated overnight at 16°C in a thermoblock. One 20  $\mu$ l control reaction was also generated, containing only the BbsI-HF<sup>®</sup>/BamHI-HF<sup>®</sup> linearized plasmid and lacking the gRNA library insert. Purification of the ligated plasmid and control were performed according to the Joung *et al.*, protocol [156]: the following day, the ligation reactions were combined with one volume 100% isopropanol, 0.01 volumes GlycoBlue<sup>™</sup> coprecipitant (Thermo Fisher), and 0.02 volumes 5M NaCl. The solution was vortexed and incubated at RT for fifteen minutes. The solution was centrifuged for fifteen minutes at RT, 15,000 xg, and the subsequent blue pellet contained the precipitated DNA. The supernatant was carefully removed, and the pellet was washed gently and without disturbance in ice-cold 80% ethanol two times, and after the ethanol was completely removed the pellet was air-dried for one minute at RT. The pellet containing the ligated plasmid was resuspended in 20  $\mu$ l of TE buffer, while the control was resuspended in 10  $\mu$ l TE buffer. The DNA solutions were incubated at 55°C for ten minutes, and the concentration was subsequently measured at the NanoDrop<sup>™</sup>.

The ligated gRNA plasmid library, the ligation control, and positive control plasmid pUC19 (New England BioLabs) were transformed into NEB Turbo electrocompetent *E. coli* (New England BioLabs) using a Gene Pulser Xcell<sup>™</sup> total electroporation system (Bio-Rad) following manufacturer instructions [232; 233] and are described including the modifications: SOC medium and 10 cm bacterial Petri dishes as well as

Corning® 245 mm square bioassay dishes (Sigma-Aldrich) containing LB agar (see Appendix 5) with 100  $\mu\text{g}/\text{ml}$  ampicillin were pre-warmed at 37°C. 1 millimeter (mm) pre-sterilized electroporation cuvettes (Molecular BioProducts) were placed on ice, and the electrocompetent *E. coli* was thawed on ice for ten minutes and gently flicked to mix. 25  $\mu\text{l}$  of electrocompetent *E. coli* were added directly to the bottom of the cuvette and gently tapped to ensure uniform distribution of the bacteria within the cuvette. 100 ng of DNA (comprising no more than 2  $\mu\text{l}$ ) was added to the bacteria, and the bacteria were electroporated with the BioRad exponential decay bacterial transformation program 1 “E. coli”, with the following specifications: voltage of 2.1 kilovolts (kV), capacitance of 200 microfarads ( $\mu\text{F}$ ), and resistance of 200  $\Omega$ . Eight total transformations of the gRNA plasmid library were performed in addition to one transformation of the ligation control reaction and one transformation of the pUC19 plasmid. 975  $\mu\text{l}$  of pre-warmed SOC was added to each cuvette immediately following transformation. The transformed bacteria were transferred to a microreaction tube and incubated at 37°C and 300 rpm for one hour in a thermoblock.

The bacteria transformed with the gRNA plasmid library were pooled, and 10  $\mu\text{l}$  of the pooled bacteria were diluted 1:1,000 in LB medium. One volume of LB medium was added to the pooled gRNA plasmid library-transformed bacteria, and 2 ml of the bacterial suspension was plated per 245 mm square bioassay dish, 100  $\mu\text{l}$  of the 1:1,000 diluted pooled gRNA plasmid library-transformed bacteria was plated on a 10 cm dish, 100  $\mu\text{l}$  of the ligation control-transformed bacteria was plated on a 10 cm dish, and 100  $\mu\text{l}$  of the pUC19-transformed bacteria was plated on a 10 cm dish. The plates were incubated at 37°C in a bacterial incubator for 14 hours. The 10 cm plate containing the 1:10,000 diluted gRNA plasmid library-transformed bacteria was used as an indicator of transformation success: the total number of colonies was counted and multiplied by 10,000 and again by 8 to determine the total number of colonies. >500 colonies per unique gRNA were necessary in order to ensure that the complexity of the gRNA library had been maintained. Additionally, 20x more total colonies were expected on the plates with gRNA plasmid library-transformed bacteria relative to the ligation control transformation.

The gRNA plasmid library-transformed bacteria was harvested from the bioassay dishes by pipetting 10 ml of LB medium onto the plate and gently scraping the colonies with a cell scraper. The bacterial suspension was collected, and the plates were washed two more times with 10 ml LB. The bacterial suspension was centrifuged

in a 50 ml tube for 10 minutes at 4°C at full speed. The supernatant was removed and the plasmid DNA was harvested using the EndoFree<sup>®</sup> Plasmid Maxi Kit (QIAGEN) following manufacturer instructions [234], with one maxiprep used per 50 ml tube. The resulting DNA solution was pooled and the DNA concentration was measured using the NanoDrop<sup>™</sup>.

The plasmid library was PCR amplified (2x NEBNext<sup>®</sup> High Fidelity PCR Master Mix (New England BioLabs), 1  $\mu$ l 20 ng/ $\mu$ l gRNA plasmid library template, 1.25  $\mu$ l 3.33  $\mu$ M TTS Universal forward primer, 1.25  $\mu$ l 3.33  $\mu$ M TTS Universal forward primer +1 stagger, 1.25  $\mu$ l 3.33  $\mu$ M TTS Universal forward primer +2 staggers, 1.25  $\mu$ l 10  $\mu$ M TTS Index 2 reverse, 19  $\mu$ l dH<sub>2</sub>O) using the following protocol:

1. 98°C for 3 minutes
2. 22 cycles of:
  - 98°C for 10 seconds
  - 63°C for 10 seconds
  - 72°C for 25 seconds
3. 72°C for 2 minutes
4. Hold at 4°C

All sequencing primers used in this work may be found in Appendix 7.10.2.2. The PCR reaction was purified using the MinElute PCR purification kit (QIAGEN) following manufacturer instructions [235]. The concentration of the PCR reaction was measured at the NanoDrop<sup>®</sup>, and 2  $\mu$ g of DNA was run in a 2% agarose gel at 100V for approximately forty minutes. The resulting 320 bp band was cut and also purified using the MinElute PCR purification kit. The plasmid library was paired-end sequenced on a MiSeq generating 18 Mio 150 bp fragments. The composition and complexity of the library was analyzed using model-based analysis of genome-wide CRISPR/Cas9 knockout (MaGECK) software [236].

### 2.9.3 Generation of lentivirus

Generation of lentivirus from the gRNA plasmid library proceeded according to the Joung *et al.* protocol [156]. In short, four 225 cm<sup>2</sup> Nunc<sup>™</sup> EasYFlask<sup>™</sup> cell culture

flasks were coated with poly-l-lysine as previously described (Section 2.1.4) and  $1.8 \times 10^7$  HEK cells were plated per flask in 45 ml D10 medium. The next day, 9 ml Opti-MEM™, 61.2  $\mu\text{g}$  lentiviral helper plasmid pMD2.G (see Appendix 7.8.5), 93.6  $\mu\text{g}$  lentiviral helper plasmid psPAX2 (see Appendix 7.8.10), and 122.4  $\mu\text{g}$  gRNA library plasmid were mixed. Plasmids pMD2.G and psPAX2 were both gifts from Didier Trono [237; 238]. Separately, 9 ml Opti-MEM™ and 1,188  $\mu\text{l}$  PLUS™ reagent (Invitrogen) were mixed, added to the plasmid mix, and incubated at RT for five minutes. In parallel, 18 ml Opti-MEM™ and 1.08 ml Lipofectamine™ 2000 were mixed and also incubated at RT for five minutes. The Lipofectamine™ and plasmid/PLUS™ mixes were combined and well mixed, and incubated at RT for five minutes further. 9 ml of the transfection mix was then added per flask of HEK cells and gently mixed, and the flasks were then incubated at 37°C in a humidified incubator for four hours. After four hours, the medium was replaced with 45 ml fresh D10 medium, and the flasks were incubated at 37°C in a humidified incubator for 48 hours. The medium was then removed and filtered with a Millex-HV 0.45  $\mu\text{M}$  polyvinylidene (PVDF) filter (Merck), aliquotted, and stored at -80°C until use.

To determine the lentiviral titer,  $1 \times 10^5$  SGET-Cas9 ESCs, which had previously been transfected with pPY-CAG-PBase and pPB-CAG-Cas9-IRES-Hygro to allow for the genomic insertion of the Cas9 gene, were plated per well of a 12-well plate pre-coated with fibronectin. In five wells, the cells were plated with 1 ml 2i+LIF medium as well as a volume of the filtered lentivirus-containing medium harvested from the HEK cell supernatant. The volumes of lentiviral supernatant added were: 10  $\mu\text{l}$ , 20  $\mu\text{l}$ , 40  $\mu\text{l}$ , 80  $\mu\text{l}$ , and 160  $\mu\text{l}$ . The medium, lentiviral supernatant, and cells were well-mixed. Two addition wells were left untreated with viral supernatant as either a negative control or a puromycin-untreated control. After 24 hours, the medium was changed to fresh 2i+LIF medium containing 12  $\mu\text{g}/\text{ml}$  puromycin, except for one of the wells without virus, which was left untreated with puromycin and received only fresh 2i+LIF medium. The medium was changed freshly every day, and after four days the cells were washed in PBS, dissociated for four minutes at 37C in TrypLE™ Express, resuspended in passaging medium, centrifuged for four minutes at 1,200 xg, RT, and finally resuspended in 1 ml 2i+LIF per condition. The cells were then counted, and the multiplicity of infection (MOI) was determined by dividing the number of cells in each lentivirus-treated well or the negative control well by the number of cells in the puromycin-untreated control, and plotting the values and



generating a linear trendline in Microsoft Excel (Microsoft). The percentage of cells infected per MOI was determined using the following equations, taken from Figliozzi *et al.*, 2016 [239]:

$$\begin{aligned}\text{Uninfected} &= e^{-MOI} \\ \text{Singly infected} &= MOI \times e^{-MOI} \\ \text{Multiply infected} &= 1 - (\text{uninfected} + \text{singly infected})\end{aligned}$$

#### 2.9.4 *Tfap2c* tiling screen in ESCs, EpiLCs, and PGCLCs

$6.43 \times 10^6$  SGET cells were plated in 25 ml 2i+LIF medium per flask of a 225 cm<sup>2</sup> Nunc™ EasYFlask™ cell culture flask pre-coated with fibronectin. Three flasks in total were used to ensure that more than  $1.54 \times 10^7$  cells were always in culture, as this is the number of gRNAs multiplied by 500 - this value is suggested by Joung *et al.* to ensure even coverage of all gRNA-containing cells throughout the screen [156]. Additionally, cells were plated at the same density in a 10 cm dish as a no-virus control. 1.684 ml lentivirus was added per flask, equivalent to an MOI of 0.3815. The cells and lentivirus were well mixed by pipetting up and down with a cannula, and the cells were incubated at 37°C in a humidified incubator for 24 hours. After 24 hours, the medium was changed to fresh 2i+LIF medium containing 12 µg/ml puromycin, and the cells were selected for twelve days in total, with the medium changed approximately every 1-2 days to fresh puromycin-containing medium as needed. The medium of the 10 cm plate was also changed to puromycin-containing medium, and was observed for cell death. The cells were passaged every 3-4 days, and approximately  $1.9 \times 10^7$  cells were always plated to ensure even gRNA coverage. After twelve days, approximately  $1.9 \times 10^7$  ESCs were induced towards EpiLCs as previously described (Section 2.1.2), and the rest of the dissociated ESCs were centrifuged and frozen in a pellet at -20°C until gDNA harvesting could occur. The EpiLC medium did not contain puromycin, and was replaced with fresh EpiLC medium after 24 hours. After 42 hours of EpiLC induction,  $1.73 \times 10^7$  EpiLCs were induced towards PGCLC fate in 23 total wells of 6-well EZSPHERE™ Microwell plates as previously described (Section 2.1.2) in puromycin-free GK15 medium. One additional well containing  $7.5 \times 10^5$  cells was maintained without cytokines, while 23 wells contained cytokines required for PGCLC induction. EpiLCs not required for PGCLC induction were centrifuged and stored as a pellet at -20°C until gDNA harvesting could occur. The EBs were maintained as previously described, and on day six of differentiation were FACS-sorted at the SH800S cell sorter for GFP+/tdTomato- cell populations

indicating correct PGCLC fate. The sorted cells were centrifuged and stored as a pellet at  $-20^{\circ}\text{C}$  until gDNA harvesting could occur. The process of ESC infection, EpiLC and PGCLC differentiation, and FACS-sorting occurred three times in total with the SGET-Cas9 cell line.

### 2.9.5 Library preparation, sequencing, and analysis of *Tfap2c* screen results

Genomic DNA from each of the cell pellets collected during the screen was isolated using the Quick-DNA™ Midiprep Plus Kit (Zymo) following manufacturer instructions [240]. The concentration of gDNA was quantified at the NanoDrop™. PCR purification of the genomically inserted gDNA sequence occurred as previously described with respect to gRNA plasmid library amplification with NEBNext high fidelity 2x PCR master mix (see Section 2.9.2), however up to  $2.5\ \mu\text{g}$  of gDNA was used as template per  $50\ \mu\text{l}$  reaction, and one of five unique reverse index primers was used per sample type in addition to the three forward stagger primers. Sequencing primers may be found in Appendix 7.10.2.2. The pooled PCR product for each condition was run on a 2% agarose gel at 100V for approximately forty minutes, cut, and gel-purified with the MinElute® gel purification kit as previously described. The previously described gRNA library plasmid amplification was also repeated. The size of each library assessed at the TapeStation prior to being sent for sequencing.

Sequencing of the libraries occurred in three runs, wherein the ESC, EpiLC, and PGCLC samples were pooled with others' libraries and sequenced in consecutive runs. All libraries were ultimately sequenced on the NovaSeq 6000 resulting in 150 bp paired-end fragments. Ultimately between 10 Mio and 27 Mio paired-end fragments were sequenced and analyzed using MaGECK-VISPR using maximum-likelihood estimation (MLE) [241], although only one single-end file was used per library for analysis. To identify positively or negatively enriched gRNAs, each condition (plasmid, ESC, EpiLC, d6 PGCLC) was compared to the previous state, *i.e.* ESC gRNA counts were compared to plasmid gRNA counts, EpiLC gRNA counts were compared against ESC gRNA counts, and PGCLC gRNA counts were compared against EpiLC gRNA counts. The reference library used to determine enrichment grouped exon-targeting GeCKO library control gRNAs, such that these control gRNAs were evaluated simultaneously to produce one enrichment score, but did not cluster individual custom gRNAs generated for the screen, which were each analyzed separately.

The scrambled gRNAs were used to normalize the gRNA read counts and determine the gRNA significance and false discovery rate (FDR) of the control and custom gRNAs. During analysis it was noted that some gRNAs designed via Benchling exhibited off-target activity, and thus all Benchling gRNAs were re-checked with the IDT CRISPR-Cas9 gRNA checker [242]. All gRNAs designed as unsuitable for use by the tool were removed from analysis, resulting in 29,192 unique test gRNAs, *i.e.* excluding GeCKO controls, scrambled gRNAs, and off-target Benchling gRNAs. Individual, significant gRNAs were annotated with the University of Santa Cruz (UCSC) Genome Browser [243] and its the Blast-like alignment tool (BLAT) [244]. Significant gRNAs and their intersections between cell types was visualized using DeepVenn [245; 246].



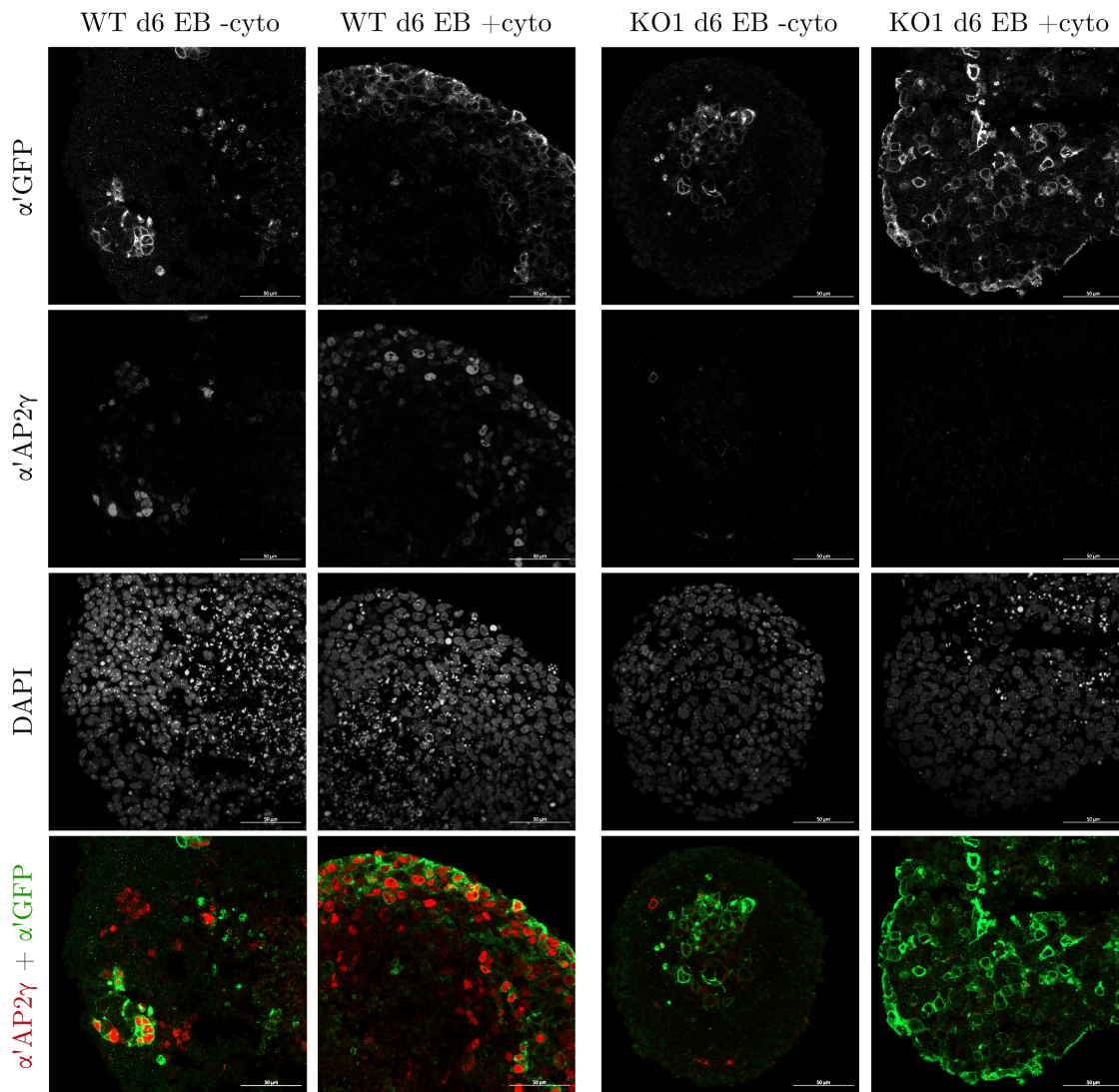
# Results

## 3.1 Knockout of AP2 $\gamma$ in mouse ESCs

To characterize the role of AP2 $\gamma$  in PGCLC differentiation, I first generated AP2 $\gamma$ -ESC lines. To achieve this, I applied CRISPR/Cas9 gene editing technology to knock AP2 $\gamma$  out of mouse ESCs by targeting Cas9 to the second exon of the gene encoding AP2 $\gamma$ , *Tfap2c*. This experiment was designed to allow for the characterization of ESCs, EpiLCs, and FACS-sorted GFP+ d4/d6 putative PGCLCs in the absence of AP2 $\gamma$ . AP2 $\gamma$ -KO lines were originally produced with Blimp1-meGFP (B1) ESCs. B1 ESCs are mouse embryonic stem cells containing a membrane-targeted enhanced GFP (meGFP) transgene which is under the transcriptional control of the regulatory elements acting on *Prdm1* [26]. *Prdm1* encodes the protein BLIMP1, which is strongly expressed in PGCs; thus, the expression of meGFP in B1 PGCLCs reports PGCLC induction *in vitro* [26].

To generate the AP2 $\gamma$ -KO ESC lines, I transfected B1 ESCs with one of two plasmids containing the Cas9 gene and a gRNA targeting exon two of *Tfap2c* (plasmids pX330-a'mTfap2c-exon2-g2 or pX330-a'mTfap2c-exon2-g3) as well as the pPY-CAG-mKO-Puro-IP plasmid containing the puromycin resistance gene encoding Puromycin N-acetyltransferase from *Streptomyces alboniger*. Following selection with puromycin for two days, I sorted single cells at the FACS. I investigated the resultant individual AP2 $\gamma$ -KO lines for the presence of mutations in exon two of *Tfap2c*. To accomplish this, I PCR-amplified the gRNA-targeted region of *Tfap2c* exon two, and cloned the amplicon into the pMiniT™ 2.0 vector. I transformed the resultant plasmid into transformation-competent *E. coli*, and sent plasmid DNA harvested from resultant single-colony inoculates for Sanger sequencing. At least ten colonies were sent for sequencing per putative AP2 $\gamma$ -KO line, to increase the likelihood that amplicons from both alleles would be sequenced. Ultimately, I identified three putative AP2 $\gamma$ -KO lines in this manner: B1 AP2 $\gamma$ -KO gRNA #3 clone #2 (here termed KO1), B1 AP2 $\gamma$ -KO gRNA #3 clone #3 (KO2), and B1 AP2 $\gamma$ -KO gRNA #3 clone #11 (KO3). Lines KO1 and KO2 contain an eight base pair (bp) deletion in exon two of *Tfap2c* which introduces a frameshift and ultimately an early stop codon (Figure 8A). One allele of line KO3 contains the same 8 bp deletion as





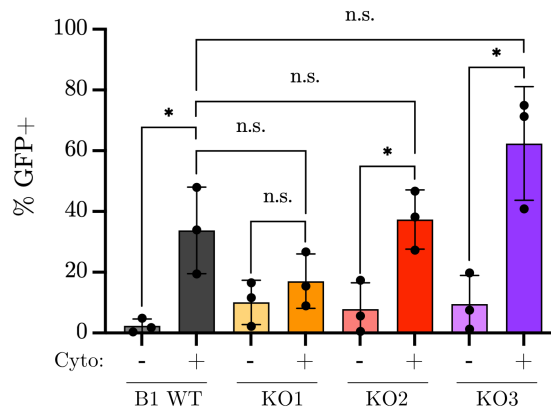
**Figure 9:** The AP2 $\gamma$ -KO phenotype is validated in IF stainings of d6 KO1 EBs. WT and KO1 EpiLCs were aggregated as EBs for six days in either -cyto control medium or +cyto medium which induces the cells towards PGCLC fate. The EBs were subsequently fixed, cryosectioned, and labeled with  $\alpha'$ AP2 $\gamma$ -Alexa-555 (shown in red),  $\alpha'$ GFP-Alexa-488 (shown in green), and DAPI. Overlay images of the  $\alpha'$ AP2 $\gamma$  and  $\alpha'$ GFP channels are shown, as are scale bars of 50  $\mu$ m.

identified in lines KO1 and KO2, however the other allele contains a one bp deletion, which nevertheless introduces a frameshift and early stop codon (Figure 8A).

I then further characterized lines KO1, KO2, and KO3 via Western blot (WB) to identify whether the introduced genomic mutations resulted in the loss of AP2 $\gamma$  on the protein level. Indeed, WB analysis demonstrated that both isoforms of AP2 $\gamma$  are completely undetectable in lysates isolated from all B1-based AP2 $\gamma$ -KO lines

(Figure 8B). I additionally performed immunofluorescence staining (IF) to further demonstrate the absence of AP2 $\gamma$  expression in the AP2 $\gamma$ -KO lines. IF staining of WT ESCs with AP2 $\gamma$  targeting-antibody demonstrates a strong signal which colocalizes with DNA-binding DAPI [184] signal, indicating that in the WT cells, AP2 $\gamma$  is localized to the nucleus as is expected for a transcription factor [133]. This strong nuclear stain is lost in all KO lines (Figure 8C). All AP2 $\gamma$ -KO lines appeared to proliferate normally in culture, although proliferation was not quantified. Lines KO1 and KO2 were observed to be generally phenotypically normal, demonstrating the typical rounded colonies characteristic of WT ESCs, although the appearance of individual cells around the edge of some colonies was noted (Figure 8D). KO3 was phenotypically distinguishable from the WT and other KO lines due to the increase in individual cells peripheral to colonies and the emergence of flatter colonies with distinguishable single cells.

I then aimed to address the question of whether the KO of AP2 $\gamma$  affects PGCLC induction. To accomplish this, I induced WT and AP2 $\gamma$ -KO ESCs first towards EpiLC fate, and subsequently towards PGCLC fate within embryoid bodies (EBs). No phenotype was observed in the AP2 $\gamma$ -KO EpiLCs relative to WT EpiLCs (data not shown), nor in the AP2 $\gamma$ -KO EBs relative to WT EBs. IF staining of d6 WT and KO1 EBs demonstrates that meGFP, reporting BLIMP1 expression, is expressed in both WT and KO1 EBs following induction with PGCLC-inducing cytokines (“+cyto” condition; Figure 9). This result demonstrates that BLIMP1 expression is not impaired by the KO of AP2 $\gamma$ . A small population of cells in the no cytokine (“-cyto”) negative control condition is also GFP+ in the WT and KO1 EBs. Strong AP2 $\gamma$  staining colocalizes with meGFP staining in WT +cyto EBs as expected, as well as



**Figure 10: GFP expression not significantly different between d6 cyto+ WT and AP2 $\gamma$ -KO cells.** The percentage of GFP+ cells indirectly reporting BLIMP1 expression in B1 WT and KO1, KO2, and KO3 EBs grown in +/- cyto medium is indicated. N = 3 for all samples. Significance of  $p < 0.05$  is indicated, and was determined using the unpaired t-test in Prism 9.



a small number of meGFP<sup>+</sup> cells in the -cyto condition. In contrast, the AP2 $\gamma$  stain is completely lost in the +/-cyto KO1 EBs, further validating that the KO of AP2 $\gamma$  was successful.

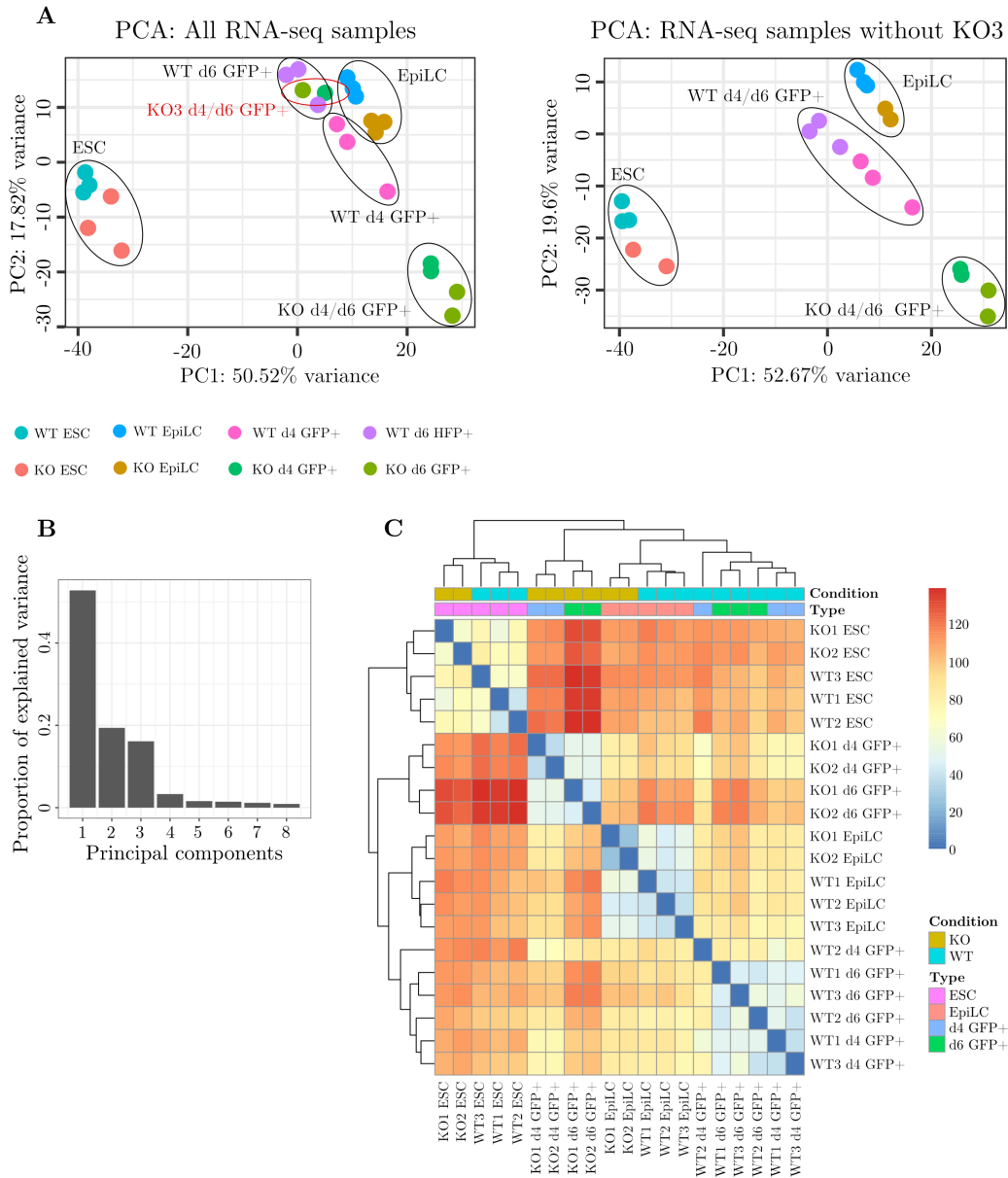
To quantify the effect of the KO of AP2 $\gamma$  on BLIMP1 expression, I FACS-sorted WT and AP2 $\gamma$ -KO cells from +/- cyto d6 EBs to determine the percentage of GFP<sup>+</sup> cells reporting BLIMP1 expression in each line (Figure 10). A significant increase in GFP<sup>+</sup> cells was detected between WT +cyto d6 cells relative to the -cyto condition, as expected. The increase in GFP<sup>+</sup> cells between the +cyto condition relative to the -cyto condition was significant in all cell lines except for KO1. The average percentage of GFP<sup>+</sup> cells varied between cells lines, which may be attributed to clonal variability. No significant difference could be detected between the percentage of GFP<sup>+</sup> WT cells and any of the KO lines, demonstrating that the KO of AP2 $\gamma$  has no effect on the expression of BLIMP1 in d6 +cyto EBs relative to WT. BLIMP1 directly induces AP2 $\gamma$  expression in PGCs [37], and as such it is unsurprising that the KO of AP2 $\gamma$  would not affect expression of its upstream regulator.

## 3.2 RNA-sequencing of AP2 $\gamma$ -KO cells

### 3.2.1 Determination of differential expression in ESCs, EpiLCs, and d4/d6 GFP<sup>+</sup> cells

To characterize the transcriptomic effect of the KO of AP2 $\gamma$  on PGCLC differentiation, I prepared RNA-sequencing (RNA-seq) libraries using WT and AP2 $\gamma$ -KO ESCs and EpiLCs, as well as d4 and d6 GFP<sup>+</sup> cells from dissociated and FACS-sorted +cyto EBs. RNA-seq allows for the direct comparison of transcriptomes between different cell types and conditions [247], and was thus performed to investigate whether the KO of AP2 $\gamma$  affected gene expression and cell fate in each collected cell type.

RNA-seq libraries were generated using 100 ng of input RNA from each sample type, and sequencing resulted in between ~36-58 million (Mio) single-end reads. I qualitatively assessed the RNA-seq libraries based on the percentage of reads which mapped to the genome, the percentage of duplicated reads, the percentage of features assigned to mapped reads, the strandedness of the data, the alignment of the reads along the gene body, and the ratio of mapped reads relative to each chromosome, and this data may be seen in Appendix 7.9.3.1.



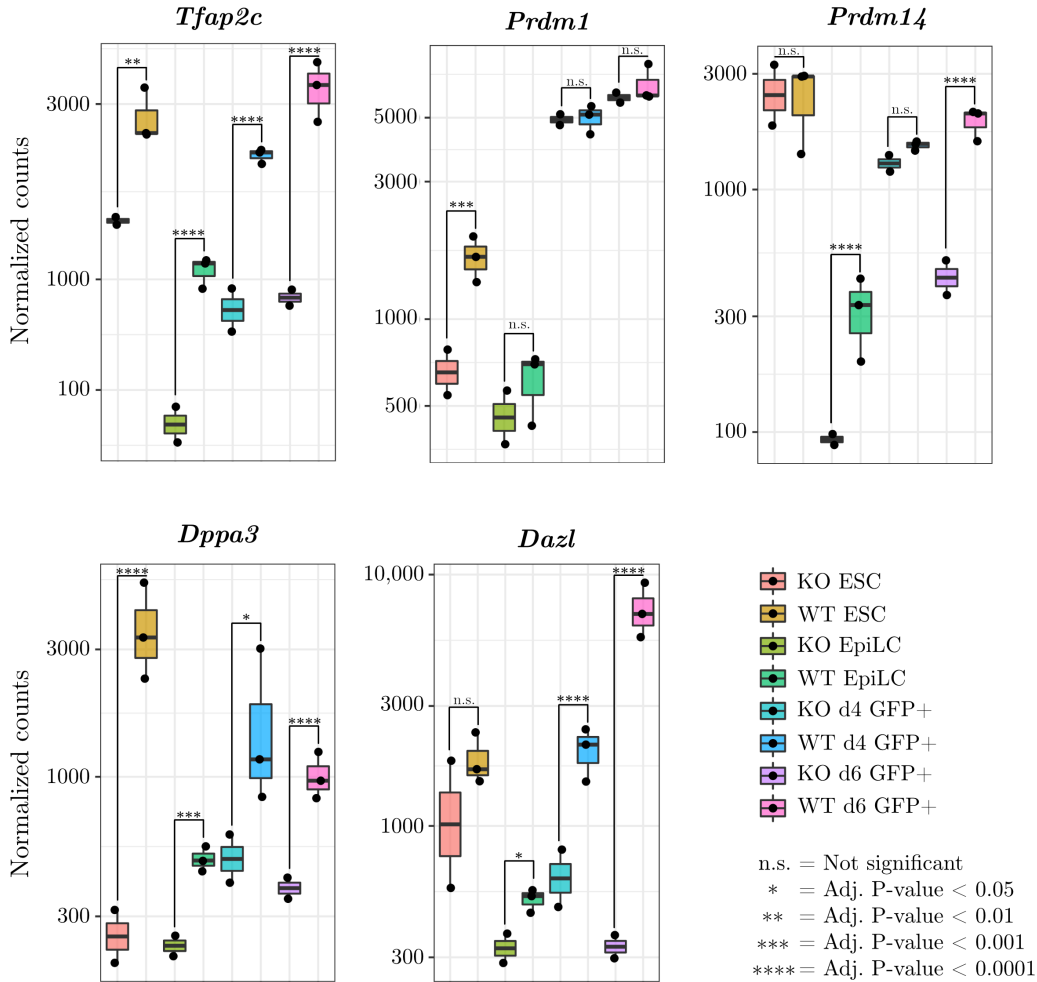
**Figure 11: RNA-seq analysis demonstrates transcriptional differences between cell types as well as WT and AP2 $\gamma$ -KO d4/d6 GFP+ cells.** A) PCA plots of RNA-seq samples with and without the inclusion of KO3 samples. KO3 d4/d6 GFP+ samples which clustered with WT d4/d6 GFP+ samples are indicated in the red ring, all other cell types and clusters are indicated in black. Color of marker indicates cell type and KO status and is explained in the key, and the variance of principal components (PC1 and PC2) is indicated. B) Scree plot of principal components used to generate PCA plots. The number of principal components and proportion of explained variance is indicated. C) Euclidian sample-to-sample distance of all RNA-seq samples, with genotype (condition) and cell type (type) indicated. Scale indicates relative difference between samples.

I performed principal component analysis (PCA) to compare the transcriptomic similarity between the samples using count tables produced from the mapped reads

(Figure 11A). Both WT and AP2 $\gamma$ -KO ESC and EpiLC samples clustered closely together. However the WT and KO1/KO2 d4 and d6 GFP+ samples clustered separately, comprising separate populations. The KO3 d4/d6 GFP+ samples clustered with the WT d4/d6 GFP+ samples, and closer examination of the KO3 lines showed measurable expression of *Tfap2c* mRNA (Appendix 7.9.3.4). Thus, the KO3 samples were removed further analysis, as they could not be considered true knockouts. The PCA scree plot is indicated to demonstrate the level of variation captured by each principal component [209], and demonstrated that most variation is accounted for in the first two dimensions (PC1 and PC2; Figure 11B). Euclidian sample-to-sample distance was also mapped to represent an unbiased comparison of sample similarity (Figure 11C) [209]. This shows that both WT and AP2 $\gamma$ -KO ESCs and WT and AP2 $\gamma$ -KO EpiLCs share a low relative distance, indicating similarity of the AP2 $\gamma$ -KO samples to the WT, while KO d4/d6 GFP+ cells demonstrate a high relative distance from the WT d4/d6 GFP+ cells, indicating sample dissimilarity.

Next, I asked whether the KO of AP2 $\gamma$  impairs the expression of genes associated with PGC fate and pluripotency. To accomplish this, I compared the expression of key pluripotency and/or PGC marker genes *Tfap2c*, *Prdm1*, *Prdm14*, *Dppa3*, and *Dazl* between samples (Figure 12). In the WT samples, all five genes are strongly upregulated in the PGCLCs relative to the EpiLCs, which is consistent with their established identities as markers of PGC fate (see Section 1.2.2). In contrast, by d4 in the AP2 $\gamma$ -KO lines *Tfap2c*, *Dppa3*, and *Dazl* are significantly downregulated, and by d6 all markers except for *Prdm1* are strongly downregulated. This result indicates that AP2 $\gamma$  is required for correct PGC factor expression, with the exception of *Prdm1*, which encodes the upstream regulator of *Tfap2c* expression BLIMP1. Interestingly, pluripotency factor *Dppa3* is also significantly downregulated in KO ESCs and EpiLCs, suggesting that AP2 $\gamma$  may play a regulatory role in the expression of some naïve pluripotency markers.

I then performed differential gene expression analysis in order to systematically analyze the transcriptomic differences between WT and AP2 $\gamma$ -KO cells. Differential gene expression was investigated between WT1-3 and KO1-2 samples per cell type with significant differential expression determined by adjusted p-value (adj. p.) < 0.01 and fold-change (fc) of +/- 1.5x. In the AP2 $\gamma$ -KO ESCs, 449 genes were differentially expressed (DE), of which 288 were upregulated and 161 were downregulated (Figure 13A). Gene ontology (GO) analysis via Goseq of DE biological process genes



**Figure 12: Key PGC markers are differentially expressed in d4/d6 GFP+ KO cells.** Expression of each gene is indicated by normalized counts (log10 scale). Significance was determined between AP2 $\gamma$ -KO and WT samples of a given cell type using DESeq2, and is indicated in the key. The color of each bar indicates the sample type, and is written in the key on the bottom right. N = 3 for WT samples, N = 2 for KO samples.

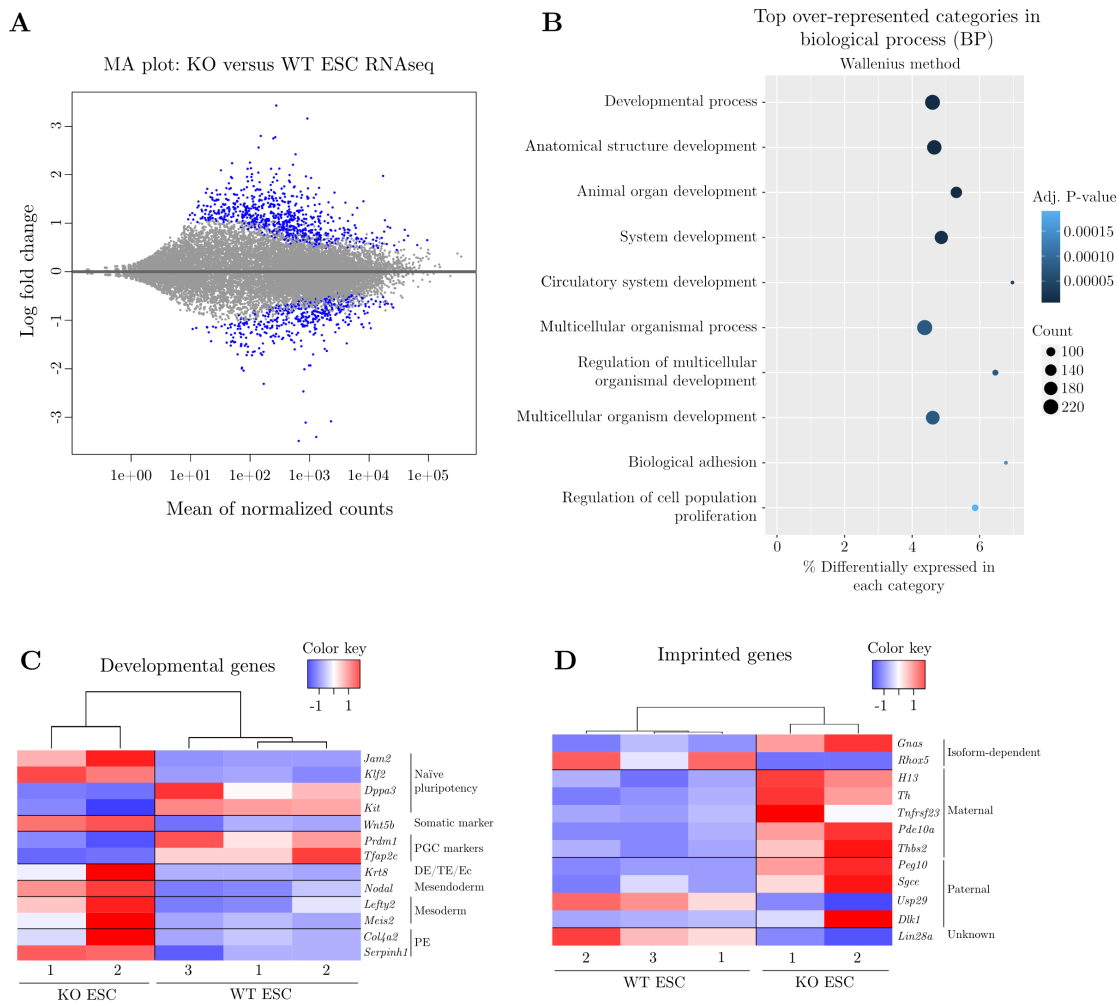
revealed many developmental terms such as “developmental process”, “system development”, and “multicellular organismal” process (Figure 13B), as well as the terms “biological adhesion” and “regulation of cell population proliferation”.

To better understand the differential expression of developmental genes in AP2 $\gamma$ -KO ESCs, I analyzed DE developmental genes of interest to this work, *e.g.* pluripotency genes, markers of specific embryonic lineages, or imprinted genes (Figure 13C). This analysis showed that naïve pluripotency markers *Dppa3* and *Kit* were downregulated while *Jam2* (Junctional adhesion molecule B) and *Klf2* (Krüppel-like factor 2) were upregulated in the AP2 $\gamma$ -KO ESCs. The PGC marker *Prdm1*, which is not strongly

expressed in WT ESCs, was downregulated even further in the AP2 $\gamma$ -KO ESCs. Markers of embryonic lineages were upregulated in the AP2 $\gamma$ -KO ESCs, including multi-lineage marker *Krt8* (Keratin 8), mesendoderm marker *Nodal* (Nodal growth generation factor), mesoderm markers *Lefty2* (Left-right determination factor 2) and *Meis2* (Meis homeobox 2), and primitive endoderm markers *Col4a2* (Collagen type IV alpha 2 chain) and *Serpinh1* (Serp family H, member 1). A small set of imprinted genes were also identified to be DE in the AP2 $\gamma$ -KO ESCs, and may be seen in Figure 13D. These findings show that modest but measurable transcriptomic misregulation occurs in the AP2 $\gamma$ -KO ESCs relative to WT, and results in the differential expression of a small number of naïve pluripotency, developmental, and imprinted genes, thus raising the question of whether AP2 $\gamma$  may play a role in the differentiation and/or pluripotency of ESCs.

I next investigated whether the AP2 $\gamma$ -KO ESCs could correctly be induced into EpiLCs. As *Tfap2c* expression is downregulated in the WT EpiLCs relative to the WT ESCs (Figure 12), *i.e.* during the exit from naïve pluripotency and into formative pluripotency, it is not expected that the KO of AP2 $\gamma$  should profoundly affect the EpiLC transcriptome. However, as the EpiLCs were induced from AP2 $\gamma$ -KO ESCs, it was of interest to determine if the differential expression identified in the AP2 $\gamma$ -KO ESCs would impair the induction towards EpiLC fate. Thus, I performed differential expression analysis and identified 792 genes which were DE between AP2 $\gamma$ -KO and WT EpiLCs, of which 436 were upregulated and 356 were downregulated (Figure 14A). GOseq analysis of all DE genes between AP2 $\gamma$ -KO and WT EpiLCs resulted in the over-representation of many developmental GO terms including “developmental process”, “anatomical structure morphogenesis”, and “cellular development process” (14B).

WT EpiLCs downregulate naïve pluripotency genes including *Dppa3* and *Prdm14* (Figure 12), while markers of formative pluripotency including *Dnmt3b* (DNA methyltransferase 3 beta), *Fgf5* (Fibroblast growth factor 5), and *Otx2* (Orthodenticle homeobox 2) are upregulated [21]. It was therefore striking that a small set of pluripotency genes were in fact downregulated even further in the AP2 $\gamma$ -KO EpiLCs relative to WT, including seven markers of naïve pluripotency and two markers of formative pluripotency, namely *Etv5* (ETS variant transcription factor 5) and *Fgf5* (Figure 14C). A small number of markers for fate-specified embryonic lineages were upregulated, including markers for the embryonic ectoderm (Ec), primitive endoderm (PE), and



**Figure 13: KO of AP2 $\gamma$  results in differential developmental gene expression in ESCs.**

A) Log ratio (M) and mean average (A) plot (MA plot) of AP2 $\gamma$ -KO versus WT ESC RNA-seq genes, generated using DESeq2. Differentially expressed (adj.  $p < 0.01$ ) genes are indicated in blue. B) Goseq gene ontology results for biological process GO terms overrepresented in DE genes between AP2 $\gamma$ -KO and WT ESC samples. Significance is indicated by the color of the marker, and gene counts associated with each cluster are indicated by the size of the marker. C) Heatmap of developmental DE genes. D) Heatmap of imprinted DE genes. All heatmaps were generated with heatmap2, and the general classification of each gene is provided to the right of the gene name, with sample name and KO status indicated below. Red indicates upregulation and blue indicates downregulation. Samples are hierarchically clustered by similarity.

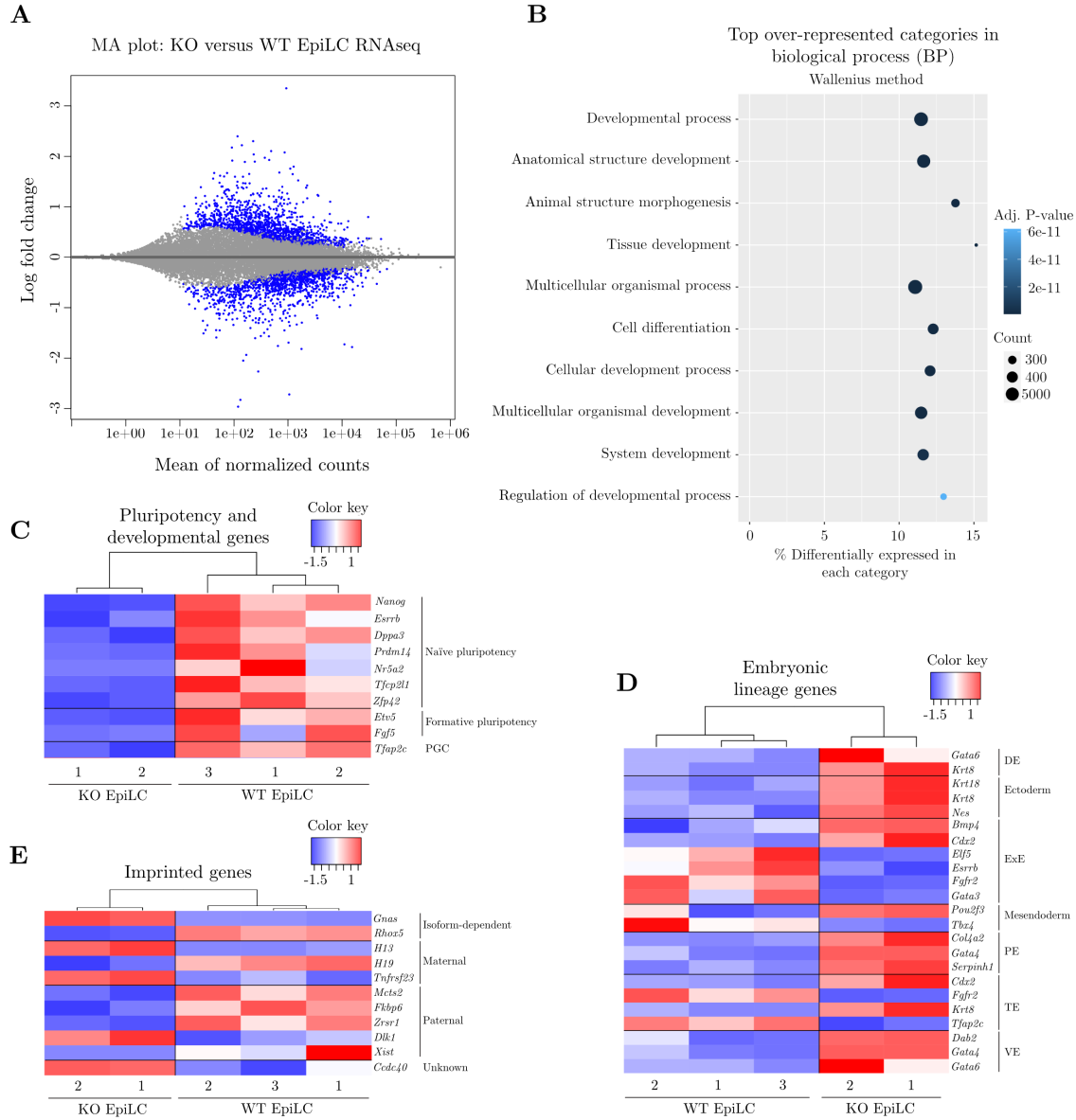
visceral endoderm (VE), while markers for the extraembryonic ectoderm (ExE) were largely downregulated (Figure 14D). Eleven imprinted genes were also identified to be DE, however these genes were both up- and downregulated in KO EpiLCs (Figure 14E). As AP2 $\gamma$  is not strongly expressed in WT EpiLCs, it is likely that the increased DE gene expression identified in the AP2 $\gamma$ -KO EpiLCs derives from misregulation originating in the AP2 $\gamma$  EpiLCs. These results suggest that AP2 $\gamma$  may be required

to correctly prime the ESCs for the exit from naïve pluripotency towards EpiLC fate.

In order to determine whether the AP2 $\gamma$ -KO EpiLCs could give rise to PGCLCs, I then investigated DE gene expression in d4 GFP+ AP2 $\gamma$ -KO cells. In total, 1,009 genes were considered significantly DE between AP2 $\gamma$ -KO and WT d4 GFP+ cells, of which 622 were upregulated and 387 downregulated (Figure 15A). GOseq GO analysis demonstrated that DE genes were enriched for terms in the developmental categories of “multicellular organism development”, “tissue development”, and “animal organ morphogenesis”, as well as the terms “biological adhesion” and “cell adhesion” (Figure 15B). However the GO term “biological adhesion”, first identified in the AP2 $\gamma$ -KO ESC GOseq analysis, is also overrepresented in the AP2 $\gamma$ -KO d4 GFP+ DE genes, as is the term “cell adhesion” (Figure 15C).

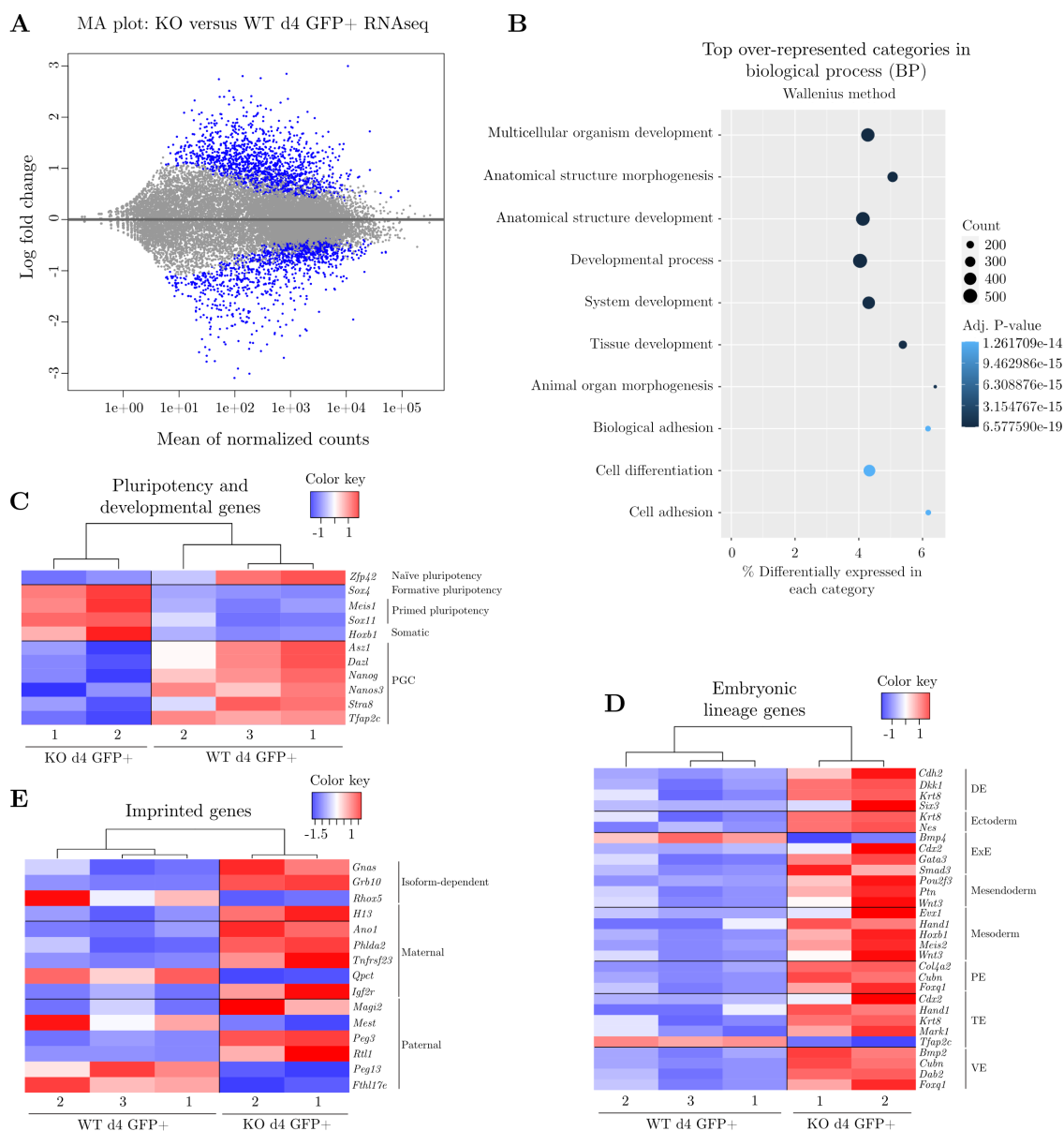
AP2 $\gamma$ -KO d4 GFP+ cells significantly downregulated six PGC factors relative to WT cells, including *Dazl* and *Nanog*. Additionally, three factors of formative and primed pluripotency were upregulated, as was the somatic marker *Hoxb1* (Homeobox B1; Figure 15C). Further, naïve pluripotency markers *Zfp42* (Zinc finger protein 42) and *Nanog* were found to be downregulated. Importantly, numerous DE genes marking non-PGC embryonic lineages were upregulated in the AP2 $\gamma$ -KO d4 GFP+ cells relative to WT, including markers for DE, mesendoderm, mesoderm, PE, TE, and VE fates (Figure 15D). Finally, fifteen imprinted genes were also found to be DE in the AP2 $\gamma$ -KO d4 GFP+ cells relative to WT (Figure 15E).

As the EBs developmentally progressed from d4 to d6, the number of DE genes dramatically increased: I identified 4,813 genes to be DE between the AP2 $\gamma$ -KO and WT d6 GFP+ cells, of which 2,644 were upregulated and 2,169 were downregulated (Figure 16A). GOseq GO analysis demonstrated that DE genes were overrepresented in developmental categories including “anatomical structure morphogenesis”, “developmental process”, “multicellular organism development”, among others, indicating that DE genes are involved in development. However, DE genes also are overrepresented in categories associated with “tube morphogenesis” and “tube development”, indicating an association with genes involved in differentiation towards somatic cell types.

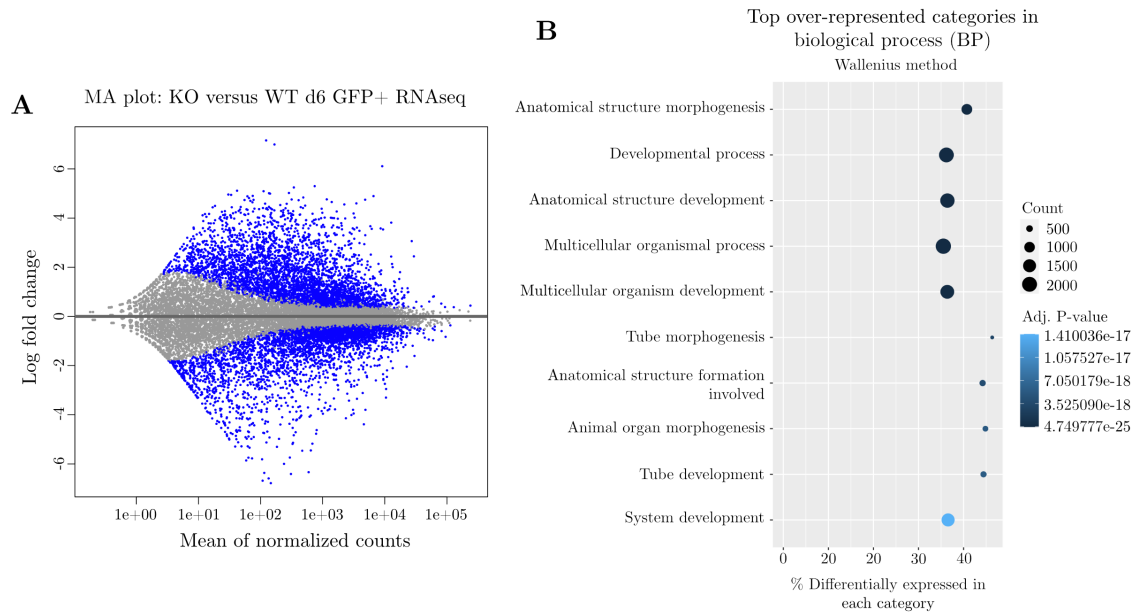


**Figure 14: KO of AP2 $\gamma$  results in differential developmental gene expression in EpiLCs.** A) MA plot of AP2 $\gamma$ -KO EpiLC versus WT RNA-seq genes, generated using DESeq2. Differentially expressed (adj.  $p < 0.01$ ) genes are indicated in blue. B) Goseq gene ontology results for biological process GO terms overrepresented in DE genes between AP2 $\gamma$ -KO and WT EpiLC samples. Significance is indicated by the color of the marker, and gene counts associated with each cluster are indicated by the size of the marker. C) Heatmap of pluripotency and developmental DE genes. D) Heatmap of imprinted DE genes. E) Heatmap of DE embryonic lineage genes. Heatmaps were generated with heatmap2, and the general classification of each gene is provided to the right of the gene name, with sample name and KO status indicated below. Red indicates upregulation and blue indicates downregulation. Samples are hierarchically clustered by similarity.





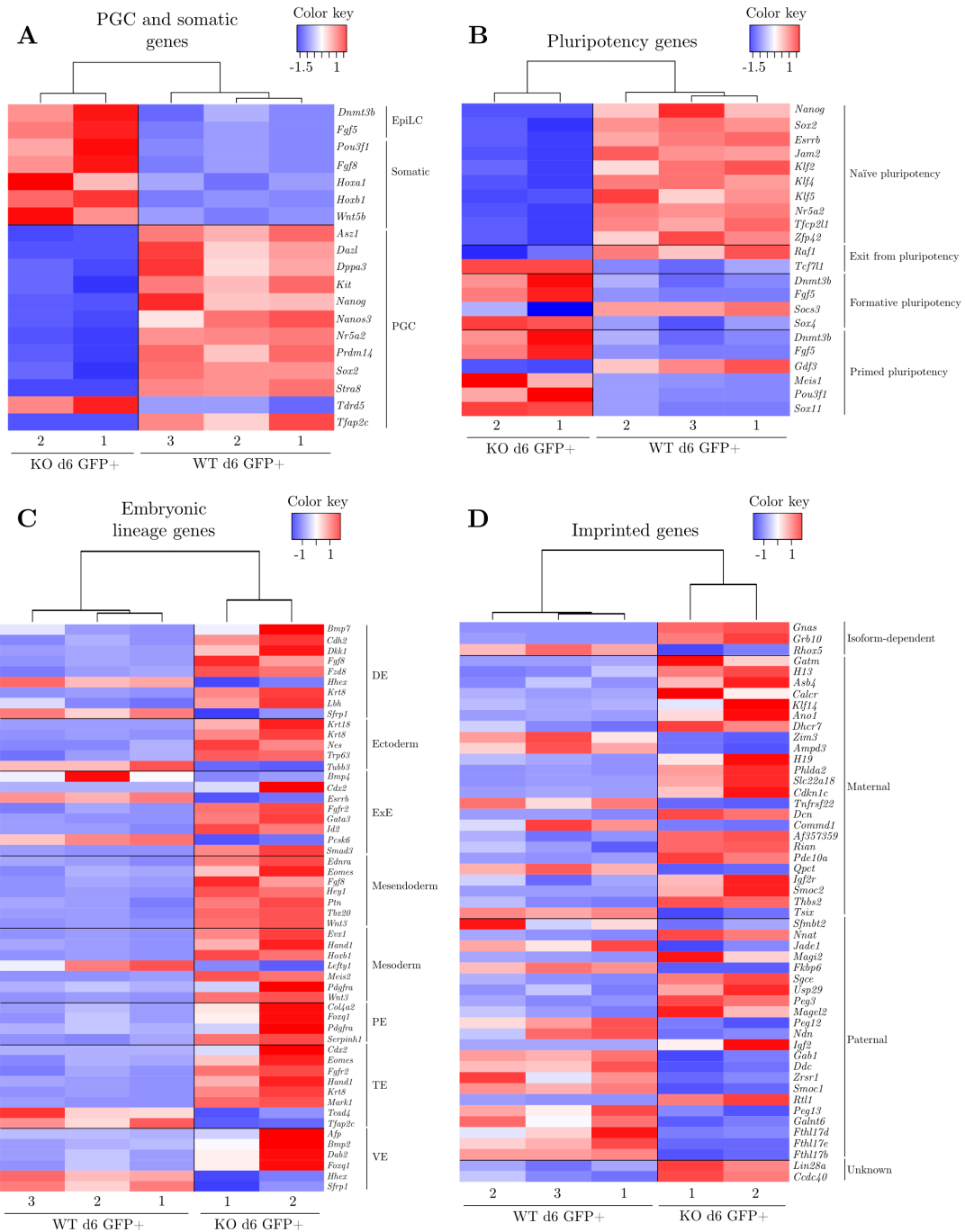
**Figure 15: KO of AP2 $\gamma$  impairs the expression of PGC and pluripotency markers in d4 GFP+ cells.** A) MA plot of AP2 $\gamma$ -KO d4 GFP+ versus WT RNA-seq genes, generated using DESeq2. Differentially expressed (adj.  $p < 0.01$ ) genes are indicated in blue. B) Goseq gene ontology results for biological process GO terms overrepresented in DE genes between AP2 $\gamma$ -KO and WT d4 GFP+ samples. Significance is indicated by the color of the marker, and gene counts associated with each cluster are indicated by the size of the marker. C) Heatmap of pluripotency and developmental DE genes. D) Heatmap of imprinted DE genes. E) Heatmap of DE embryonic lineage genes. Heatmaps were generated with heatmap2, and the general classification of each gene is provided to the right of the gene name, with sample name and KO status indicated below. Red indicates upregulation and blue indicates downregulation. Samples are hierarchically clustered by similarity.



**Figure 16: KO of AP2 $\gamma$  profoundly alters differential expression of developmental genes in d6 GFP+ cells.** A) MA plot of AP2 $\gamma$ -KO d6 GFP+ versus WT RNA-seq genes, generated using DESeq2. Differentially expressed (adj. p. < 0.01) genes are indicated in blue. B) GOseq gene ontology results for biological process GO terms overrepresented in DE genes between AP2 $\gamma$ -KO and WT d6 GFP+ samples. Significance is indicated by the color of the marker, and gene counts associated with each cluster are indicated by the size of the marker.

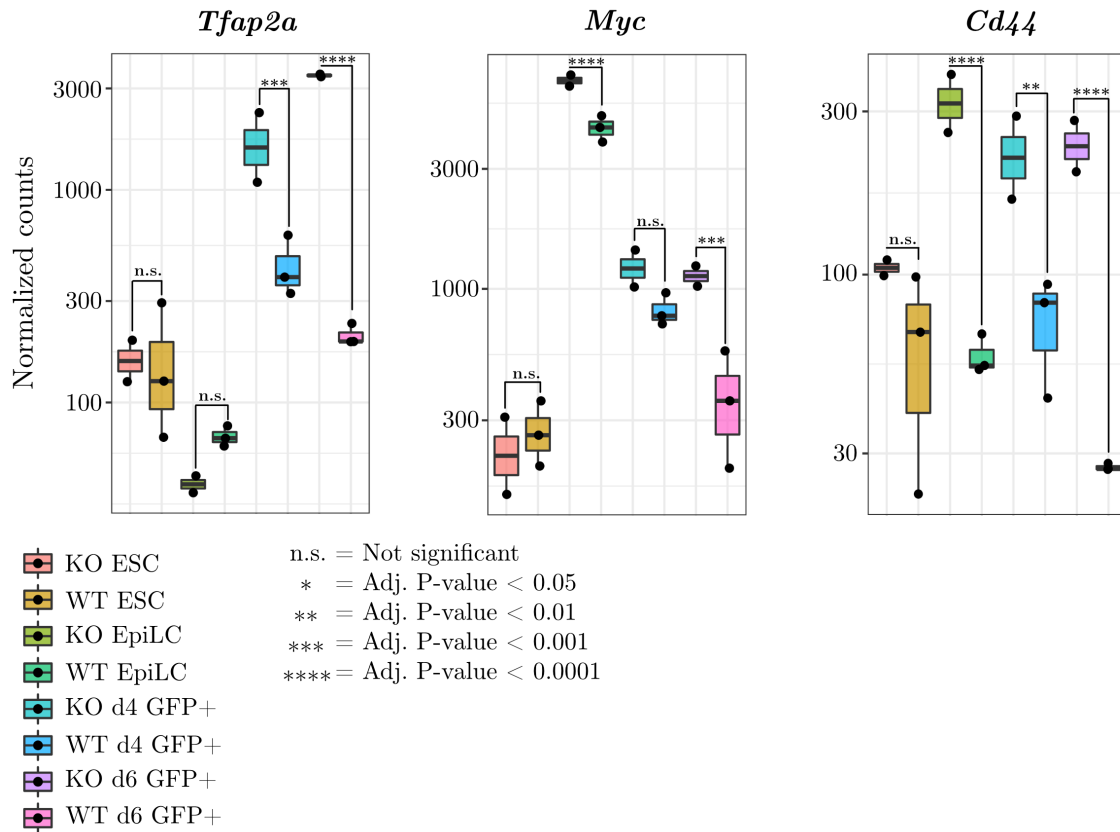
The KO of AP2 $\gamma$  profoundly affects the expression of DE genes in the d6 GFP+ cells; many known PGC and pluripotency markers were downregulated in the KO cells relative to the WT, including *Dppa3*, *Prdm14*, and *Sox2*, while markers of formative and primed pluripotency were upregulated, including *Dnmt3b* and *Fgf5* (Figure 17A-B). Further, numerous markers of non-PGC embryonic lineages, including DE, Ec, ExE, mesendodermal, mesodermal, PE, TE, and VE fates were upregulated in the AP2 $\gamma$ -KO cells relative to WT (Figure 17C). Additionally, the number of imprinted genes which were determined to be DE in the d6 AP2 $\gamma$ -KO GFP+ cells increased relative to d4 (Figure 17D).

Overall, the analysis of DE genes in the AP2 $\gamma$ -KO d6 GFP+ cells clearly indicates that PGCLC differentiation is impaired, as PGC and naïve pluripotency markers are uniformly downregulated and somatic markers as well as formative and primed pluripotency markers are correspondingly upregulated. Further, the aberrant expression of many markers of somatic or extraembryonic lineages shows global transcriptional misregulation, and reinforces the non-PGCLC identity of the AP2 $\gamma$ -KO cells. Additionally, numerous imprinted genes were identified to be DE in the AP2 $\gamma$ -KO d6 GFP+



**Figure 17: Many pluripotency, somatic, and imprinted genes are aberrantly expressed in AP2 $\gamma$ -KO d6 GFP $^{+}$  cells relative to WT.** A) Heatmap of PGC and somatic fate DE genes. B) Heatmap of DE pluripotency genes. C) Heatmap of DE embryonic lineage genes. D) Heatmap of DE imprinting genes. Heatmaps were generated with heatmap2, and the general classification of each gene is provided to the right of the gene name, with sample name and KO status indicated below. Red indicates upregulation and blue indicates downregulation. Samples are hierarchically clustered by similarity.

cells; as imprinting is regulated by DNA methylation [248], this raises the possibility that AP2 $\gamma$  is associated with epigenetic regulation during PGCLC differentiation. Finally, it should be noted that many more genes were identified as DE in the AP2 $\gamma$ -KO d6 GFP+ cells relative to d4, which demonstrates that the effect of the KO of AP2 $\gamma$  increases with developmental time during PGCLC differentiation.



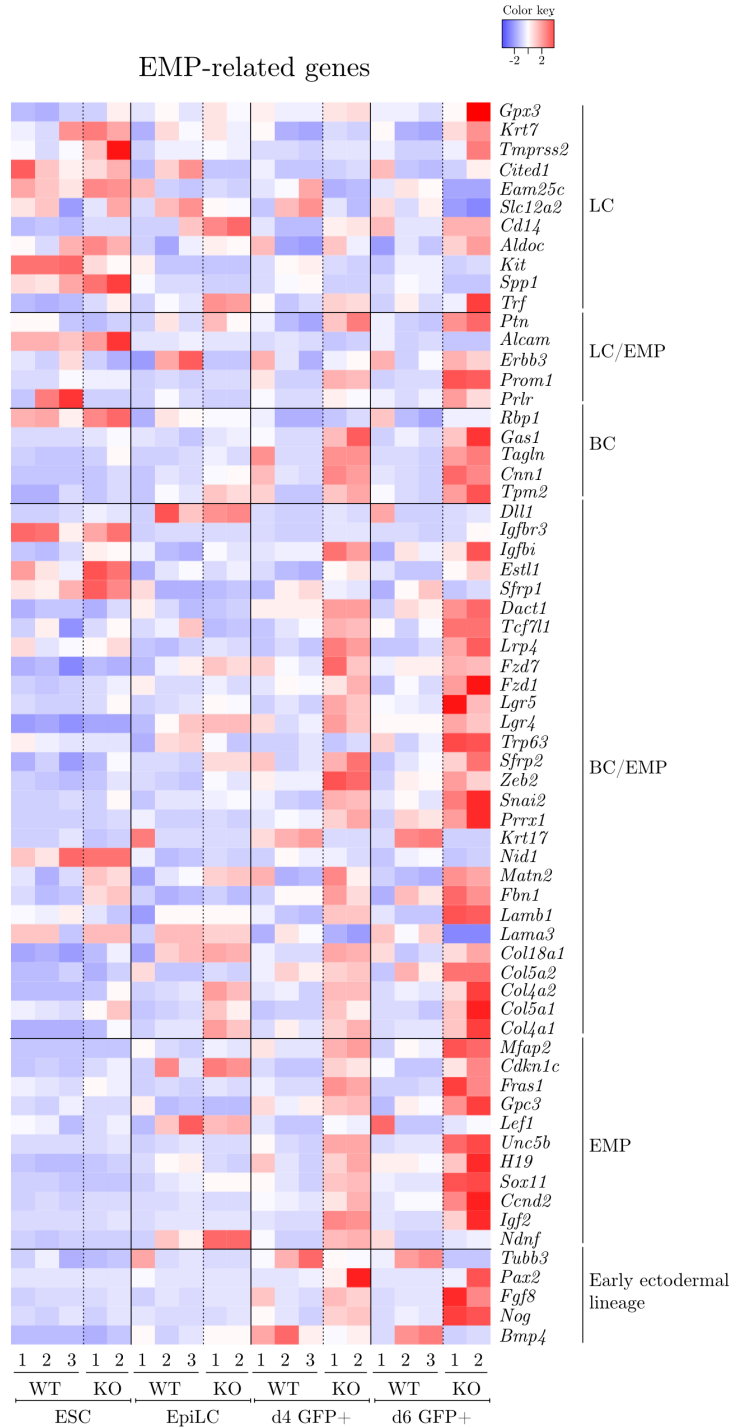
**Figure 18:** *Tfap2a*, *Myc*, and *Cd44* are DE in AP2 $\alpha$ -KO d6 GFP+ cells. The normalized counts of DE genes of interest *Tfap2a*, *Myc*, and *Cd44* in AP2 $\alpha$ -KO and WT ESCs, EpiLCs, and PGCLCs are shown. Significance between AP2 $\alpha$ -KO and WT samples was determined with DESeq2. WT N = 3, AP2 $\alpha$ -KO N = 2. Normalized counts are shown (log10 scale), and the sample/significance key is included below the plots.

It should be noted that during the examination of the RNA-seq data, I identified three particular DE genes of interest: *Tfap2a*, *Myc*, and *Cd44* (CD44 antigen) (Figure 18). *Tfap2a* encodes AP2 $\alpha$ , the protein within the AP2 family most closely associated with AP2 $\gamma$  (discussed in Section 1.3.1), and is upregulated in d4 and d6 GFP+ AP2 $\alpha$ -KO cells relative to WT cells. This fact is intriguing, as the potential redundancy of AP2 $\gamma$  and AP2 $\alpha$  has been discussed in literature [88; 106; 107]. *Myc* is a well-characterized oncogene [249] and is upregulated in AP2 $\alpha$ -KO EpiLCs and d6 GFP+

cells relative to WT, while *Cd44* is a marker of cancer stem cells and is upregulated in AP2 $\gamma$ -KO EpiLCs and d4/d6 GFP+ cells relative to WT [250]. As AP2 $\gamma$  has been implicated in many forms of cancer, both as an ameliorating or exacerbating factor (see Section 1.3.1), it is thus interesting that the AP2 $\gamma$ -KO EpiLCs and d4/d6 GFP+ cells differentially express known cancer markers.

### 3.2.2 Upregulation of EMP markers in AP2 $\alpha$ -KO d4/d6 GFP+ cells

The upregulation of numerous non-PGC lineage factors and the downregulation of pluripotency markers in AP2 $\alpha$ -KO d4/d6 GFP+ cells raised the question of whether this was the result of uniform misregulation of gene expression, or whether a program of differentiation towards a specific fate was occurring in the absence of AP2 $\alpha$  and under the experimental conditions of PGCLC induction, *i.e.* in the presence of BMP4 or other cytokines. To address this, I investigated published works describing embryonic cell populations characterized by BLIMP1 expression, to identify whether the described cells also expressed genes identified as upregulated in the d6 GFP+ KO line. BLIMP1 is expressed in highly proliferative cells of the pubescent mammary gland [251], and AP2 $\gamma$  has been identified as a regulator of post-natal mammary development [131]; when knocked-out in mammary tissue, the branching of mammary epithelia is impaired although not lost [131] and cells are skewed towards a basal rather than luminal fate [76]. Thus, I specifically surveyed published works concerning embryonic mammary specification. A 2018 paper by Wuidart *et al.* describes a precursor embryonic mammary gland population termed embryonic multipotent precursors (EMPs) [252]. These cells emerge by approximately E14 in the mouse embryo, and express markers including *Sox9* (SRY-box transcription factor 9) and *Sox11* (SRY-box transcription factor 11), and do not express the E18.5 mammary tissue marker *Sox10* (SRY-box transcription factor 10). Similarly, *Sox9* and *Sox11* are both significantly upregulated (adj. p. < 0.05, fc +/- 1.5x) in the AP2 $\alpha$ -KO d6 GFP+ cells, while *Sox10* is not upregulated and is lowly expressed (<18 normalized counts in any d6 WT/AP2 $\alpha$ -KO samples).



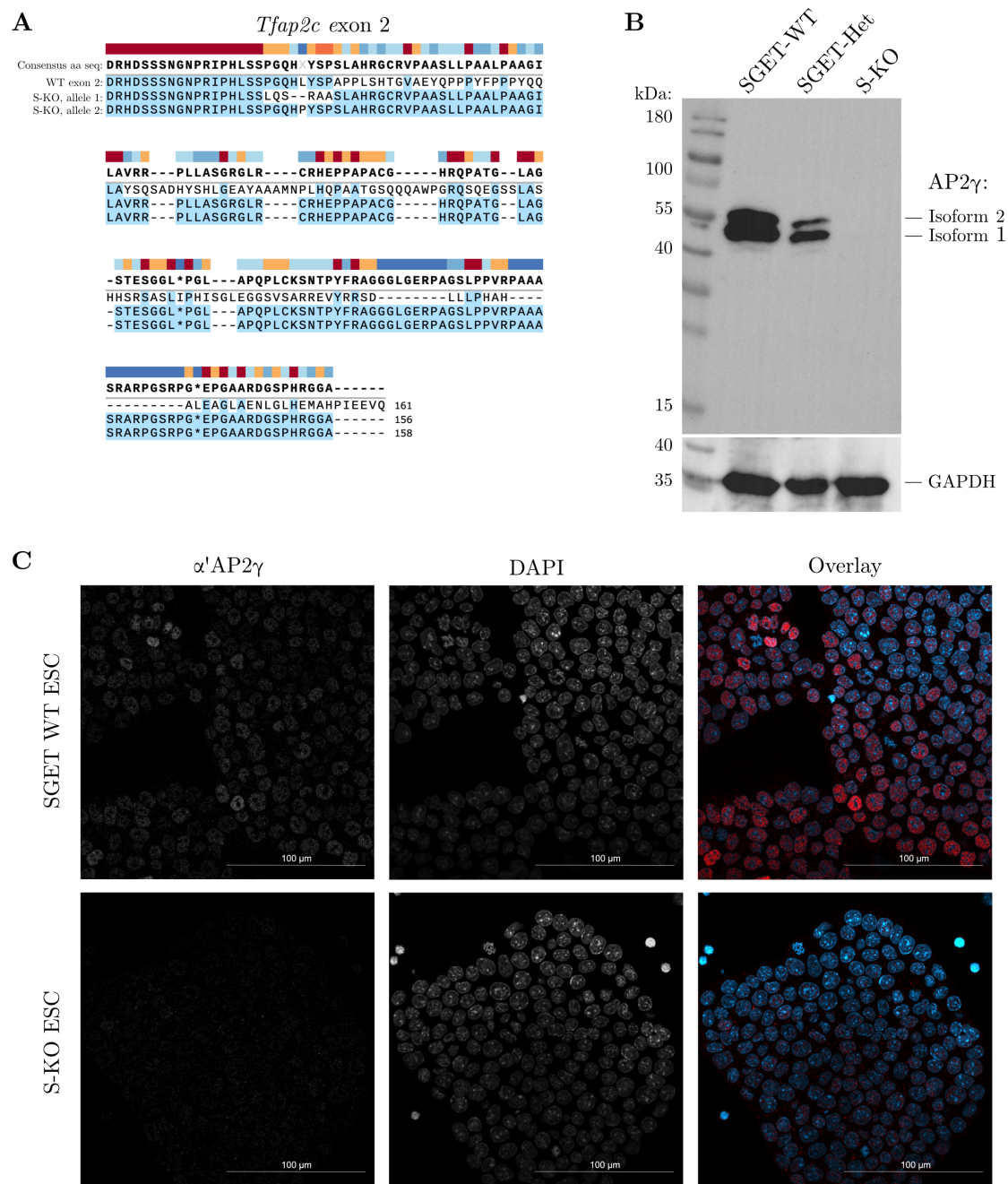
**Figure 19: AP2 $\gamma$ -KO d4/d6 GFP<sup>+</sup> cells upregulate many markers of EMP fate.** EMP gene expression was examined by determining the significance (adj.  $p$ . < 0.01,  $fc$  +/- 1.5x) of DE genes in AP2 $\alpha$ -KO d6 GFP<sup>+</sup> cells against WT, and using heatmap2 to visualize the expression of these genes as the cells develop from ESCs towards d6 GFP<sup>+</sup> cells. LC indicates luminal cell, EMP indicates embryonic multipotent progenitor cell, and BC indicates basal cell. LC/EMP and BC/EMP indicate markers present in both cell populations. The general classification of each gene is provided to the right of the gene name. Red indicates upregulation and blue indicates downregulation.

The expression of genes upregulated in EMPs as well as adult luminal (LC) and basal (BC) mammary cell populations, as identified through single-cell sequencing in the (scRNA-seq) Wuidart *et al.* study, was examined in the AP2 $\alpha$ -KO d6 GFP+ cells from this study. Numerous genes identified as markers of EMP cells, markers shared by both BC and EMP (BC/EMP) cells, and BC markers were also upregulated in the AP2 $\alpha$ -KO d6 GFP+ cells relative to WT cells (Figure 19). However, few markers of LC fate or markers shared by LC and EMP (LC/EMP) cells were upregulated in the AP2 $\alpha$ -KO d6 GFP+ cells. Overall, the upregulation of EMP markers underlines the divergence of the AP2 $\gamma$ -KO d4/d6 GFP+ cells from the PGC gene program and towards a somatic fate, with a bias towards specific cell types.

### 3.2.3 Validation of the AP2 $\gamma$ -KO phenotype in an SGET ESC line

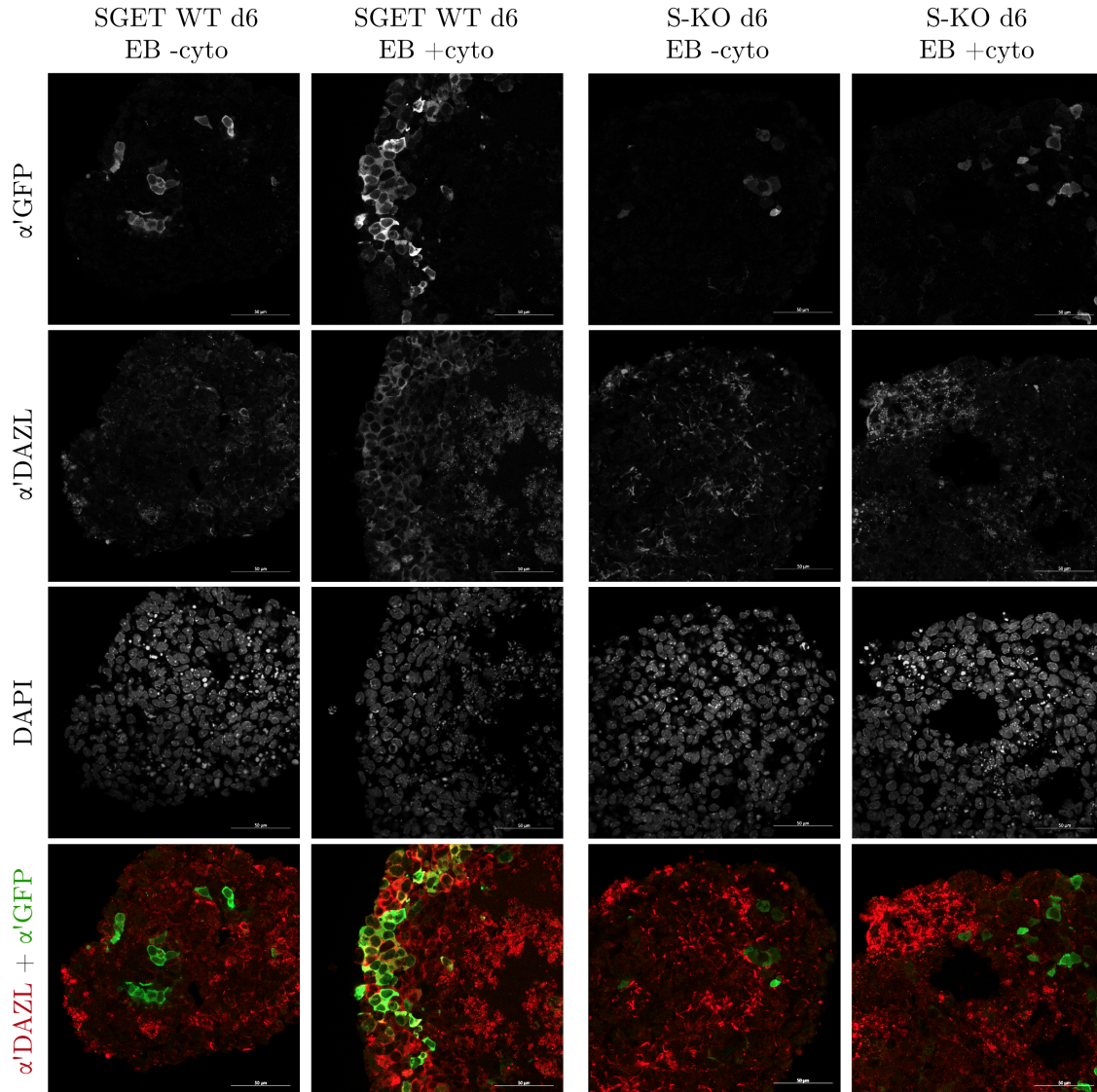
To validate and confirm my previous results from the AP2 $\gamma$ -KO B1-based lines, I developed an Stella-eGFP ESG1-tdTomato (SGET) ESC line harboring the KO of AP2 $\gamma$ . SGET ESCs contain an eGFP gene located upstream of the stop codon of the gene *Dppa3*, encoding Stella, as well as a *Tdtomato* gene in-frame of exon one of an allele of the gene *Dppa5a* (Developmental pluripotency associated 5), encoding ESG1 [172]. This cell line is a “traffic-light” reporter, wherein the strong expression of Stella in ESCs and PGCLCs and the strong expression of ESG1 in ESCs and EpiLCs allows for the fluorescent identification of three discrete cell populations; the GFP+/tdTomato+ ESCs, the GFP-/tdTomato+ EpiLCs, and the GFP+/tdTomato- PGCLCs [172].

I generated the AP2 $\gamma$ -KO SGET line by targeting exon two of *Tfap2c* as previously described (Section 3.1). The generated SGET AP2 $\gamma$ -KO gRNA #2 clone #13 (S-KO) ESC line was validated via Sanger sequencing to have separate frame-shifting mutations on each allele, exhibiting an 8 bp or 14 bp deletion, respectively. Both mutant alleles introduce an early stop codon (Figure 20A). Western blotting (Figure 20B) and IF staining (Figure 20C) demonstrated complete depletion of both isoforms of the AP2 $\gamma$  protein in the S-KO line. Another potential AP2 $\gamma$ -KO line was characterized, here termed S-Het (Figure 20B), and appeared to demonstrate a heterozygous KO phenotype and was thus excluded from further analysis.



**Figure 20: Validation of SGET-based AP2 $\gamma$ -KO ESC line S-KO.** A) Comparison of translated AP2 $\gamma$  protein sequences between SGET-WT ESCs and AP2 $\gamma$ -KO ESC line S-KO. Premature stop codons are indicated as asterisks (\*), and blue-highlighting demonstrates consensus between the WT and AP2 $\gamma$ -KO aa sequences. B) Western blot of SGET WT, S-Het, and S-KO ESC lysates, with 80  $\mu$ g protein/lane. AP2 $\gamma$  isoforms 1 and 2 are indicated, with expected size in kDa of 49 and 53 kDa, respectively. The upper blot indicates  $\alpha'$ AP2 $\gamma$ -HRP labeling, while the lower blot indicates sequential  $\alpha'$ GAPDH-HRP labeling of the same blot, with expected size of 37 kDa. A protein ladder indicates sizes in kDa in the leftmost lane. C) 40x confocal images depicting the IF staining of SGET WT and S-KO ESCs.  $\alpha'$ AP2 $\gamma$ -Alexa-647 is shown in red and DAPI is shown in blue, with scale bars of 100  $\mu$ m indicated.





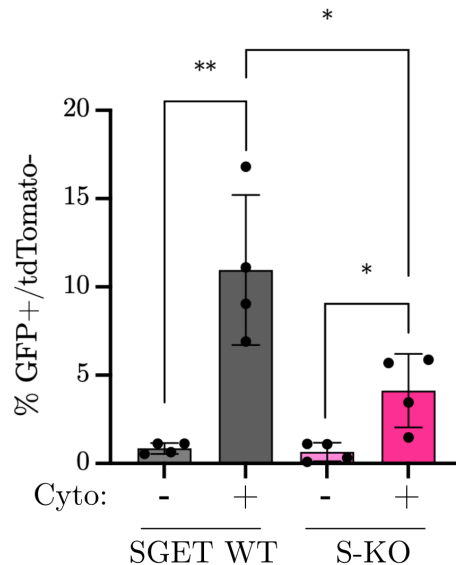
**Figure 21: KO of AP2 $\gamma$  impairs PGC marker expression on d6 of PGCLC differentiation in S-KO cells.** WT SGET and S-KO EpiLCs were aggregated as EBs for six days in either -cyto control medium or +cyto medium which induces the cells towards PGCLC fate. The EBs were subsequently fixed, cryosectioned, and labeled with  $\alpha'$ GFP-Alexa-488 (shown in green),  $\alpha'$ DAZL-Alexa-647 (shown in red), and DAPI. Overlay images of the  $\alpha'$ DAZL and  $\alpha'$ GFP channels are shown, as are scale bars indicating 50  $\mu$ m.

To determine whether the effects of the KO of AP2 $\gamma$  identified in B1 cells were maintained in a distinct fluorescent reporter line, I induced SGET WT and S-KO cells towards PGCLC fate and fixed the resulting EBs on d6 of differentiation for cryosectioning and IF staining to analyze the expression of PGC markers. The sections were labeled with primary and fluorescence-coupled secondary antibodies targeting AP2 $\gamma$  and DAZL for confocal imaging. The SGET WT d6

+cyto cells demonstrate strong GFP expression (reporting Stella expression) as well as strong cytosolic DAZL expression, with many of the Stella+ cells colocalizing with DAZL+ cells, indicating mature PGC fate (Figure 21). In contrast, few S-KO d6 +cyto cells were GFP+ or DAZL+, and cells which were positive for either label rarely colocalized. This result is consistent with RNA-seq data from the B1 AP2 $\gamma$ -KO lines demonstrating downregulated *Dazl* and *Dppa3* expression on d6 of differentiation (Figure 12).

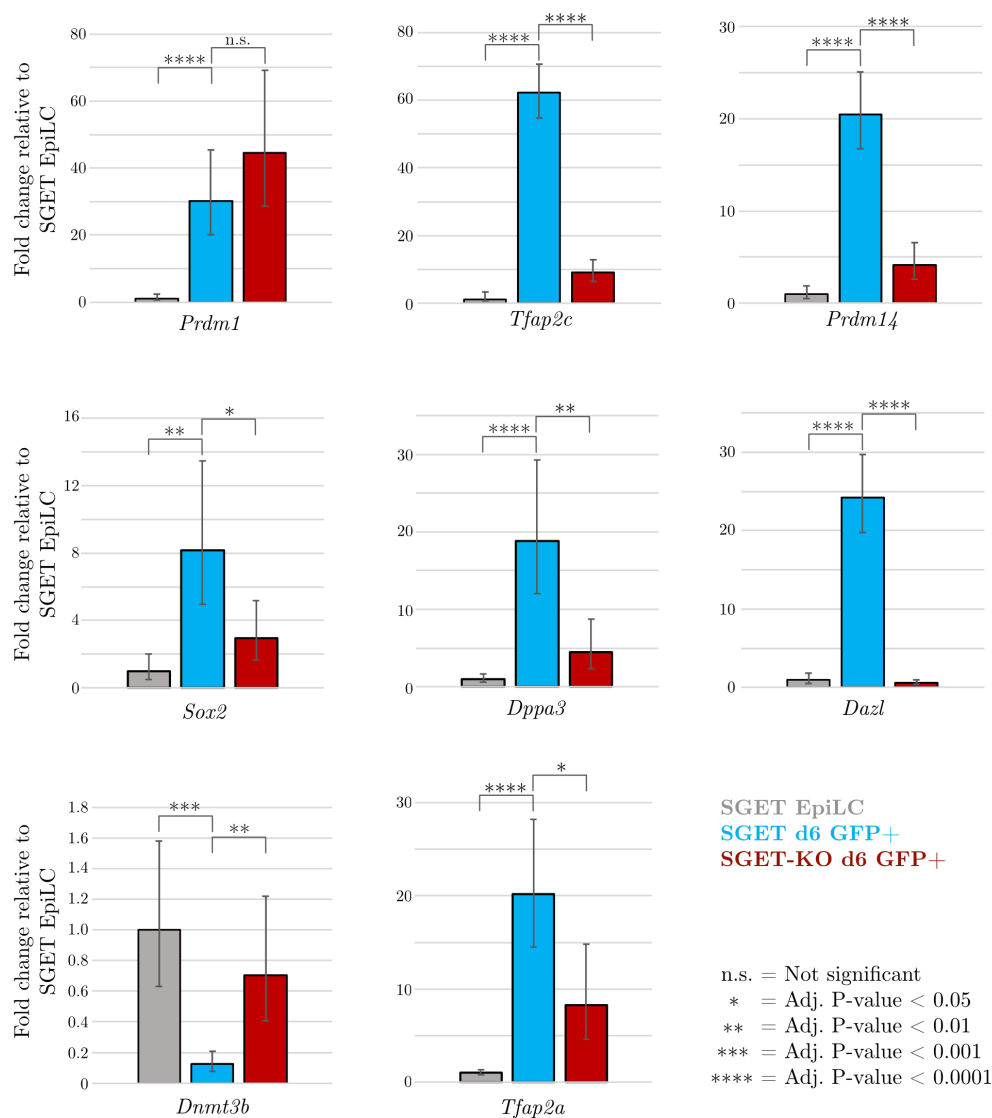
To quantify the number of cells reporting PGCLC fate in the AP2 $\gamma$ -KO context, d6 +/- cyto SGET WT and S-KO EBs were dissociated and FACS-sorted. As per the traffic-light fluorescence profiles of the SGET cells, d6 GFP+/tdTomato- cells indicate correct PGCLC induction [172]. SGET WT d6 +cyto EBs generally gave rise to ~11% GFP+/tdTomato- cells, while S-KO d6 +cyto EBs gave rise to far fewer GFP+/tdTomato- cells on average, at ~4% (Figure 22). This data further reinforces the finding that *Dppa3* expression is downregulated in the absence of AP2 $\gamma$  12).

To characterize the gene expression profiles of the S-KO cells, I isolated RNA from FACS-sorted GFP+/tdTomato- cells from SGET WT and S-KO d6 +cyto EBs. I subsequently performed qRT-PCR to investigate whether DE genes identified by RNA-seq were similarly up- or downregulated in the S-KO GFP+/tdTomato- cells (Figure 23). Consistent with my previous data (Figures 10, 12, 17A), *Prdm1* expression was not downregulated in the S-KO d6 GFP+/tdTomato- cells relative to the SGET WT, while downregulation of *Tfap2c*, *Prdm14*, *Sox*, and *Dazl* expression and upregulation of *Dnmt3b* expression was indeed observed. However, *Tfap2a* expression is in fact downregulated in the S-KO d6



**Figure 22: Loss of GFP+/tdTomato-cells in the d6 +cyto EBs S-KO line indicates impaired PGCLC differentiation.** The percentage of GFP+/tdTomato- cells reporting Stella+/ESG1- expression in SGET WT and S-KO EBs grown in +/- cyto medium for six days is indicated. N = 3 for all samples. Significance (p < 0.05) is indicated and was determined using the unpaired t-test in Prism 9.

GFP<sup>+</sup>/tdTomato<sup>-</sup> cells relative to SGET WT, although upregulation of *Tfap2a* was observed in the AP2 $\gamma$ -KO d6 GFP<sup>+</sup> RNA-seq data (Figure 18). This suggests that while the S-KO line largely confirms that while the downregulation of PGC factors in d6 AP2 $\gamma$ -KO cells is AP2 $\gamma$ -specific and not a cell-line specific effect, it appears that the d6 Blimp1 GFP<sup>+</sup> and SGET GFP<sup>+</sup>/tdTomato<sup>-</sup> cells may represent different populations to some extent.



**Figure 23: S-KO d6 GFP<sup>+</sup>/tdTomato<sup>-</sup> cells do not upregulate PGC markers.** The qRT-PCR analysis of SGET EpiLC (grey bars), SGET d6 GFP<sup>+</sup>/tdTomato<sup>-</sup> cell (blue bars), and S-KO d6 GFP<sup>+</sup>/tdTomato<sup>-</sup> cell cDNA (red bars) for genes determined to be DE in d6 B1-based AP2 $\gamma$ -KO GFP<sup>+</sup> cells is shown. Fold change is indicated relative to SGET-EpiLC, and standard deviation is indicated via error bars. Significance of  $p < 0.05$  was determined using the z-test in Microsoft Excel.  $N = 3$  for all samples.

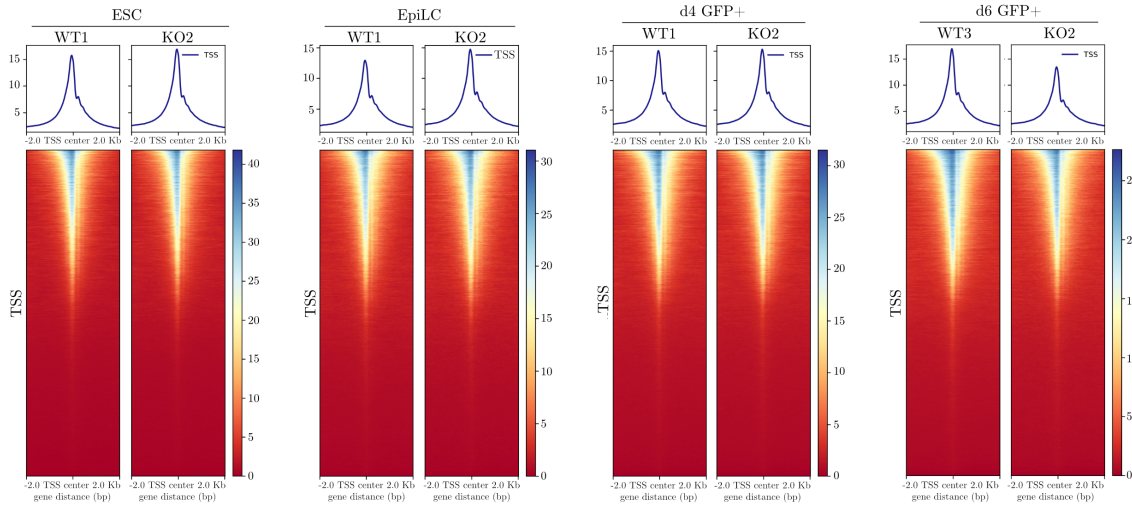
### 3.3 Analysis of chromatin accessibility in AP2 $\gamma$ -KO cells

#### 3.3.1 TSS accessibility examined in AP2 $\gamma$ -KO cells

RNA-seq analysis demonstrated that in the absence of AP2 $\gamma$ , the transcriptome of every examined cell type is altered. As AP2 $\gamma$  has been described as a potential pioneer factor [73; 77; 52], *i.e.* a TF capable of opening closed chromatin [253], I wanted to investigate whether the transcriptomic misregulation identified in the AP2 $\gamma$ -KO cells was related to altered chromatin accessibility. Thus, I prepared and analyzed ATAC-seq libraries from B1-based WT and AP2 $\gamma$ -KO ESCs, EpiLCs, and d4/d6 GFP+ cells. 50,000 ESCs and EpiLCs, and 25,000 thousand d4/d6 GFP+ cells were collected from WT and KO1/2 B1-based lines for ATAC-seq library processing, amplification, and purification. The libraries were sequenced to produce between  $\sim$ 23-85 Mio fragments per 50 bp paired-end library. The libraries were qualitatively assessed based on per-sequence GC content, sequencing quality scores, sequence length pre- and post-adaptor trimming, percentage genomic alignment, and percentage read duplication (Appendix 7.9.1.1), as well as fragment length distribution (Appendix 7.9.1.2).

To examine if the KO of AP2 $\gamma$  had a global effect upon chromatin accessibility, I examined the ATAC-seq read alignment to all transcriptional start sites (TSSs) in each library using the deepTools computeMatrix tool [213]. The TSS of a gene represents the location where transcription of a gene is initiated [254], and is located downstream of the promoter and within the nucleosome-free region (NFR) [255]. Downstream of the TSS, the characteristic “nucleosomal ladder” indicating nucleosomes wrapped with DNA may be detected [254]. The ATAC-seq libraries were examined to determine if there was a global TSS profile phenotype upon depletion of AP2 $\gamma$  which resulted in an altered TSS profile. To accomplish this, the libraries were aligned +/- 2 kb around known TSS coordinates to allow for the visual representation of promoter and nucleosomal ladder enrichment. In both the WT and AP2 $\gamma$ -KO samples, a clear peak is detected at the TSS, indicating the accessible NFR has been detected (Figure 24). Additionally, a faint nucleosomal ladder may be detected immediately adjacent to the center of the TSS in all samples. Thus, the KO of AP2 $\gamma$  does not appear to affect global TSS accessibility nor the presence of the nucleosomal ladder downstream of the promoter sequence. The mapped TSS

profiles for all libraries may be found in Appendix 7.9.1.3.

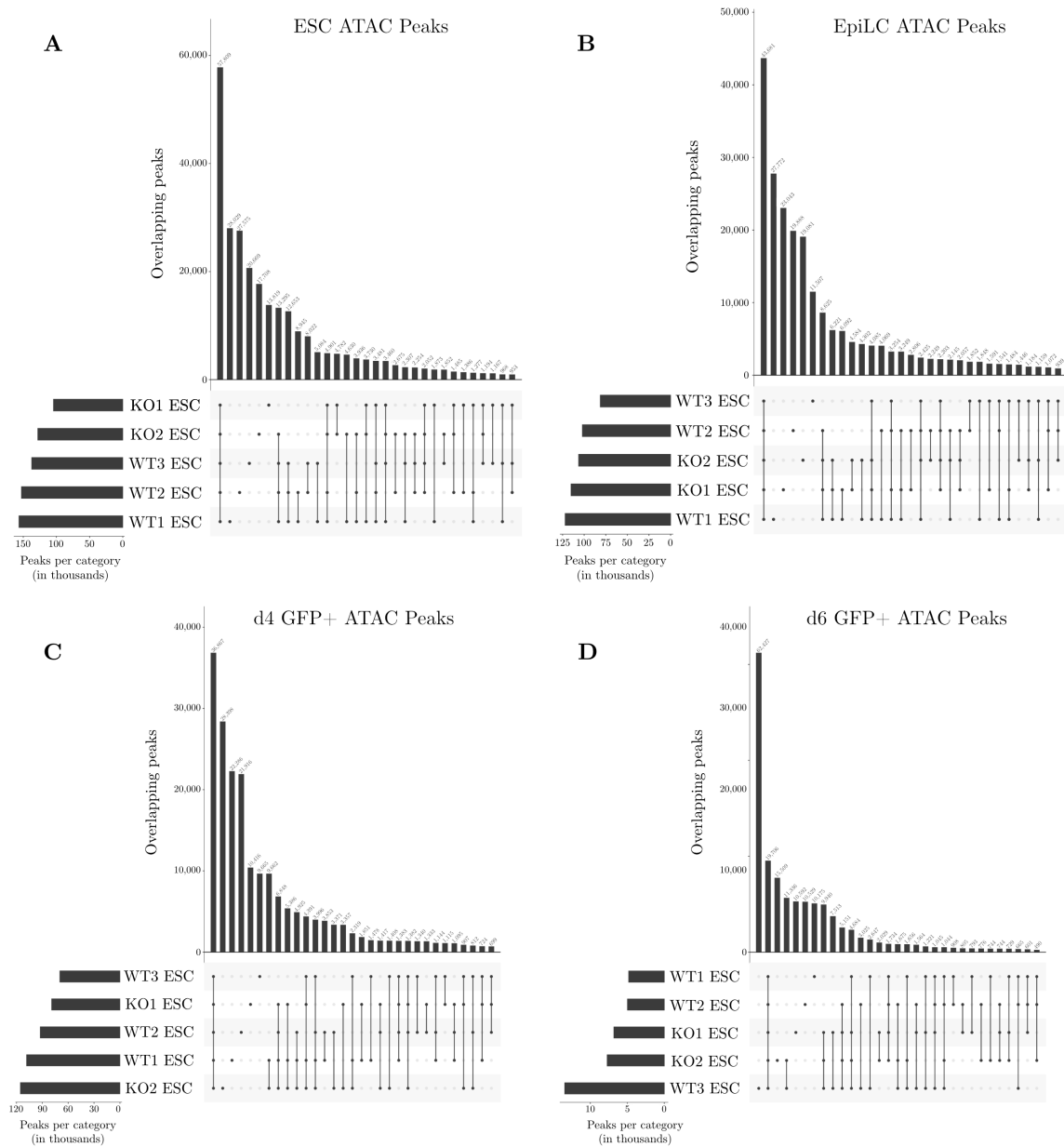


**Figure 24: The TSS chromatin accessibility profile is not altered in AP2 $\gamma$ -KO cells.** The alignment of mapped reads at transcription start sites (TSS) of ATAC-seq libraries is shown as a heatmap, with one example of a WT or AP2 $\gamma$ -KO sample per cell type and KO status. The cell type and specific cell line used are both indicated, as are the number of reads in thousands (0-15), and the distance from the TSS in bp. Color indicates reads aligning to the TSS, with blue indicating high alignment and red indicating low alignment. The average TSS profile per library is shown above the heatmap.

### 3.3.2 Characterization of DAPs upon AP2 $\gamma$ -KO

In order to identify and analyze genomic coordinates significantly enriched for aligned reads, I called peaks with a false discovery rate (FDR)  $< 0.05$  on ATAC-seq libraries using MACS2. The number of unique and shared peaks between AP2 $\gamma$ -KO and WT samples within a particular cell type was compared using the HOMER function `getDifferentialPeaks` and plotted using the R package `UpSetR` (Figure 25). In ESC, EpiLC, and d4 GFP+ samples, the majority of peaks are shared between all of the WT and AP2 $\gamma$ -KO samples of a cell type. However, the WT3 d6 GFP+ sample showed a large number of unique peaks relative to the other d6 samples (Figure 25D). This dissimilarity was confirmed by performing PCA analysis (Appendix 7.9.1.4), which demonstrated that the d6 GFP+ sample WT3 clusters separately from the other samples. In addition, this sample also has a more discernible nucleosomal ladder (Appendix 7.9.1.2). Thus, it is likely that the WT3 d6 GFP+ sample differs from WT1-2 d6 GFP+ samples due to the number of callable peaks and the resulting genomic resolution this data provides. This information should be taken into account

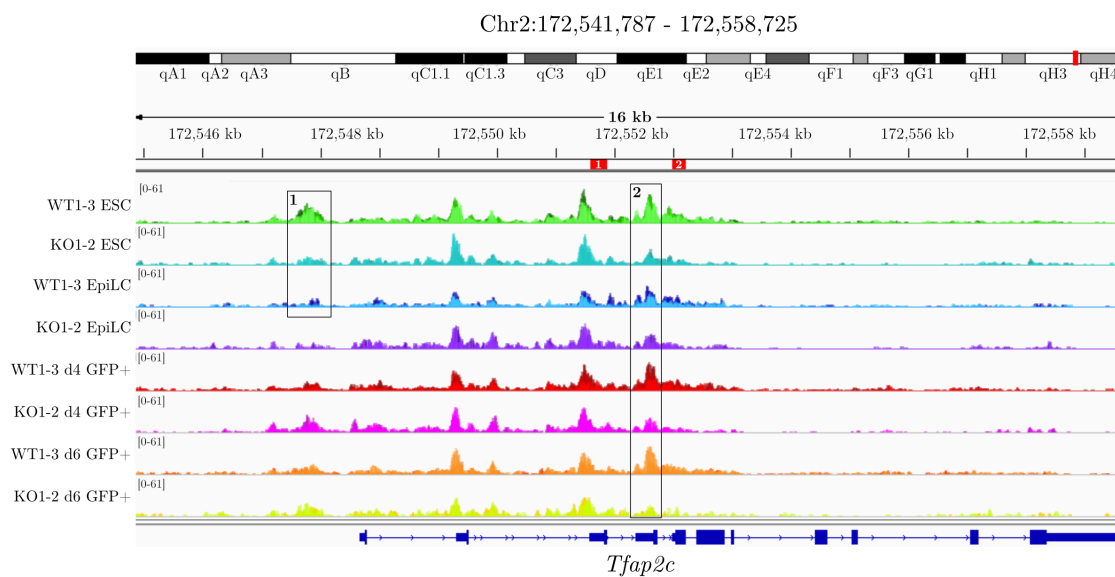
when assessing the d6 GFP+ results.



**Figure 25: Overlapping and unique peaks in ATAC-seq libraries.** The number of unique and overlapping peaks in each ATAC-seq library was determined using the HOMER function `getDifferentialPeaks`, and plotted using UpSetR. The number of total peaks per sample is indicated, as are the number of unique and shared peaks per individual sample or per multiple samples. A) ESC peaks. B) EpiLC peaks. C) D4 GFP+ peaks. D) D6 GFP+ peaks.

Towards the aim of identifying and characterizing the genomic regions which demonstrated altered chromatin accessibility upon the KO of AP2 $\gamma$ , I identified differentially accessible peaks (DAPs) using EdgeR (significance FDR < 0.01, fc +/- 2x).

DAPs are defined in this work as peaks which may be positively enriched (a new peak is called or an existing peak increases in size) or negatively enriched (a peak is lost or an existing peak decreases in size) when two sample sets are compared. Figure 26 visually demonstrates example DAPs which were identified at the *Tfap2c* gene locus, and are differential between WT cell types as well as between WT and AP2 $\gamma$ -KO samples. I then began the process of DAP characterization by performing motif enrichment analysis using the HOMER genomic known TF motif enrichment tool findMotifsGenome.pl. Motif enrichment analysis allows for the identification of TF DNA-binding motifs in genomic sequences, and thus for the potential correspondence of identified TF motifs with TF activity in a particular sample [256].



**Figure 26: Interactive genome viewer (IGV) tracks of WT and AP2 $\gamma$ -KO ATAC-seq libraries with DAPs indicated at the *Tfap2c* gene locus.** WT1-3 and KO1-2 Bigwig files demonstrating genomically aligned reads are overlaid by cell type and AP2 $\gamma$ -KO status. Peak 1 indicates a peak that is differentially present in WT ESCs but is lost in AP2 $\gamma$ -KO ESCs and WT EpiLCs. Peak 2 indicates a peak that is present in all WT samples, however is reduced in size in AP2 $\gamma$ -KO ESCs and d4/d6 GFP<sup>+</sup> cells. Chromosomal coordinates and increments are indicated. Transcription start sites for *Tfap2c* isoforms 1 and 2 are indicated in red adjacent to the chromosomal coordinates, with isoform 1 (indicated by “1”) located upstream of isoform 2 (indicated by “2”).

I first investigated motifs enriched in DAPs between WT cell states, as the enrichment of naïve or primed pluripotency TF motifs within DAPs of the correct cell type may act as a quality control to ensure that peak-calling and motif enrichment was performed correctly and that the cells accurately represent the correct cell type. In this context, naïve-specific peaks refers to DAPs found to be significantly enriched in



WT ESCs relative to WT EpiLCs, while primed-specific peaks refer to DAPs found to be significantly enriched in WT EpiLCs relative to WT ESCs. An additional reason to investigate the motif enrichment of the WT dataset is the fact that the AP2 $\gamma$  DNA-binding motif was found to be enriched in naïve-specific ATAC-seq peaks of both mouse and human [257], so it was of interest to determine whether the same effect could be identified in the ATAC-seq data analyzed in this work.

The motif enrichment analysis of DAPs identified substantial differences between WT cell states. 29.3% of all peaks were considered differentially accessible in WT EpiLCs relative to WT ESCs, with 15.2% (23,801/156,620) of peaks negatively and 9.4% (14,723/156,620) positively enriched (Appendix 7.9.1.5A). 28.9% of peaks were found to be differentially accessible in WT d4 GFP+ PGCLCs relative to WT EpiLCs, with 6.2% (6,726/108,866) negatively and 4.4% (4,789/108,866) positively enriched (Appendix 7.9.1.5B). Only one DAP was found between WT d6 and d4 GFP+ PGCLCs; the negatively enriched peak is located at chr14:24,828,865-24,829,445 within a predicted CRE. I then investigated naïve-specific DAPs for motif enrichment, and both AP2 $\gamma$  and AP2 $\alpha$  were found to be negatively enriched in EpiLC DAPs relative to ESCs, *i.e.* accessible genomic regions containing AP2 $\gamma$  and AP2 $\alpha$  motifs became less accessible once the cells had differentiated from ESCs to EpiLCs (Figure 27A). These naïve-specific peaks also demonstrated the upregulation of motifs associated with naïve pluripotency factors SOX2, KLF4, OCT4, NANOG, and NR5A2 (Nuclear receptor subfamily 5 group A member 2), among others. Examination of primed-specific peaks demonstrated the enrichment of motifs associated with primed pluripotency, including OCT6 (Octamer transcription factor 6, encoded by *Pou3f1*) and OTX2, as well as formative pluripotency factor SOX4 (SRY-box transcription factor 4), in addition to naïve pluripotency factors SOX2 and OCT4 (Figure 27B). Interestingly, regulators of three-dimensional (3D) genome architecture CTCF (CCCTC-binding factor) and CTCFL (CCCTC-binding factor-like; also known as BORIS) [258] are also positively enriched in the primed-specific peaks. It should finally be noted that two negatively enriched DAPs were identified on enhancers identified to be acting on known enhancers of the gene *Pou5f1*, and may be seen in Appendix 7.9.1.6. Overall, this analysis indicates that motifs of naïve and primed pluripotency TFs are indeed enriched in the correct cell type, and confirms that motifs of AP2 $\gamma$  as well as AP2 $\alpha$  are enriched in the naïve-specific peaks. Next, I investigated motifs negatively enriched in the d4 GFP+ PGCLCs relative to



A Negatively enriched motifs in DE peaks (EdgeR WT EpiLC-ESC)				B Positively enriched motifs in DE peaks (EdgeR WT EpiLC-ESC)			
Rank:	Motif:	Name:	P-Value:	Rank:	Motif:	Name:	P-Value:
1		KLF5	1e-1229	1		OCT4/SOX2/ TCF/NANOG	1e-285
3		SOX2	1e-1066	5		CTCF	1e-199
4		KLF4	1e-1010	7		CTCF	1e-174
10		OCT4/SOX2/ TCF/NANOG	1e-830	11		OCT4	1e-122
14		ESRRB	1e-717	17		SOX2	1e-92
20		TFCP2L1	1e-571	66		OTX2	1e-20
33		OCT4	1e-311	101		KLF4	1e-10
36		NR5A2	1e-300	103		KLF5	1e-9
42		AP2 $\alpha$	1e-236				
44		AP2 $\gamma$	1e-221				
118		NANOG	1e-30				

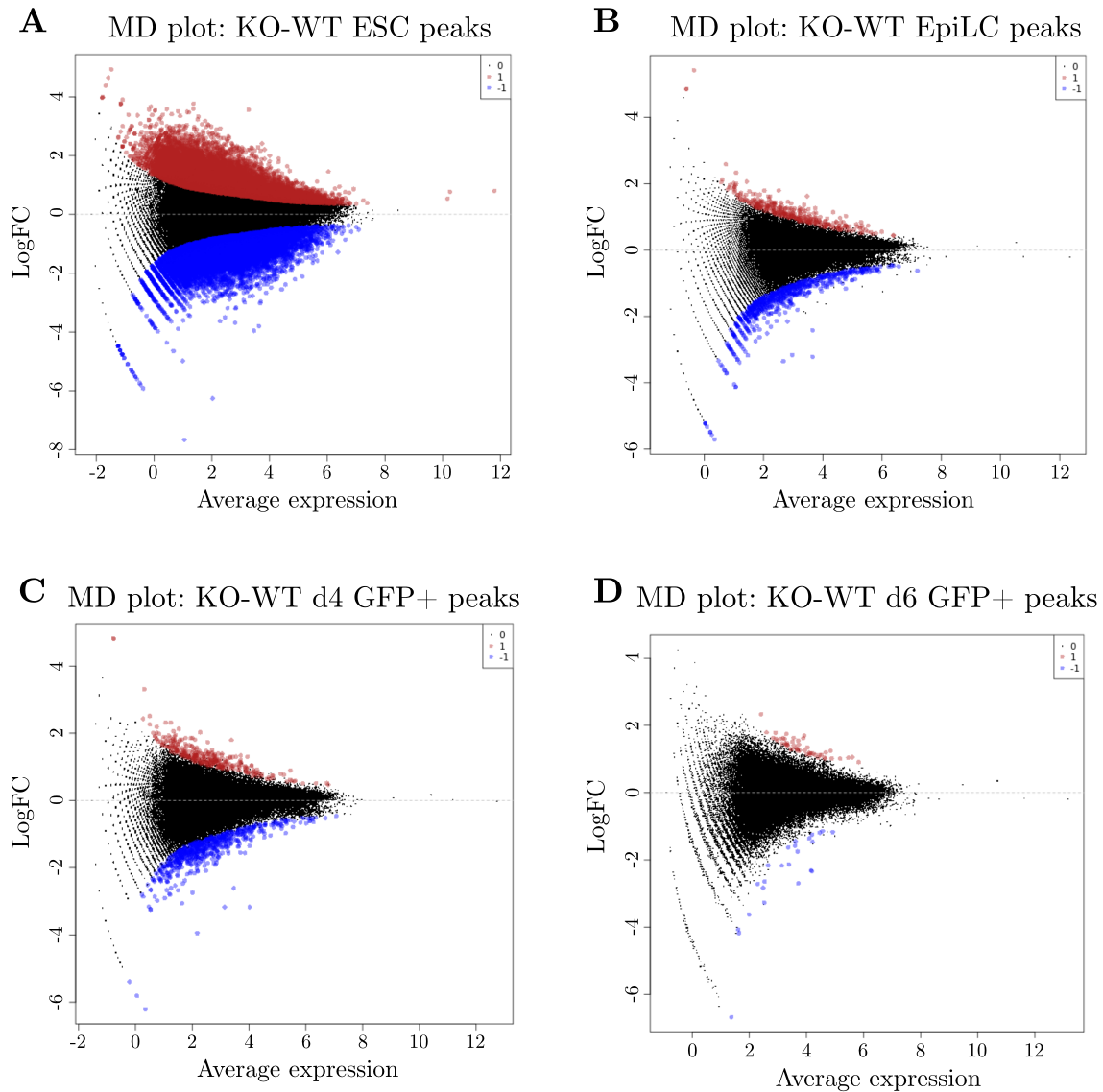
C Negatively enriched motifs in DE peaks (EdgeR WT d4 PGCLC-EpiLC)				D Positively enriched motifs in DE peaks (EdgeR WT d4 PGCLC-EpiLC)			
Rank:	Motif:	Name:	P-Value:	Rank:	Motif:	Name:	P-Value:
1		OCT4/SOX2/ TCF/NANOG	1e-389	1		AP2 $\alpha$	1e-247
5		OCT4	1e-165	2		AP2 $\gamma$	1e-211
15		SOX2	1e-96	3		SOX2	1e-196
128		BRACHYURY	1e-6	33		OCT4/SOX2/ TCF/NANOG	1e-73
152		OTX2	1e-4	75		OCT4	1e-29
167		CTCF	1e-3	92		NANOG	1e-21
				116		ESRRB	1e-15
				135		CTCF	1e-10
				211		BLIMP1	1e-3

**Figure 27: AP2 $\gamma$  and AP2 $\alpha$  are enriched in naïve- and PGC-specific ATAC-seq peaks.** Motif enrichment analysis of DAPs negatively or positively enriched in WT cells compared to the previous cell state. The rank and p-value calculated by HOMER are indicated, and significance is determined by p-value < 1e-3. A) Negatively enriched motifs in WT EpiLCs relative to ESCs. B) Positively enriched motifs in WT EpiLCs relative to ESCs. C) Negatively enriched motifs in d4 GFP+ PGCLCs relative to EpiLCs. D) Positively enriched motifs in d4 GFP+ PGCLCs relative to EpiLCs.

EpiLCs, *i.e.* accessible genomic regions which became less accessible once the cells had differentiated from EpiLCs towards PGCLC fate. Negatively enriched motifs included motifs for primed pluripotency factor OTX2 as well as Myc, a factor identified to be upregulated in EpiLCs relative to d4/d6 GFP+ PGCLCs (Figure 18). Additionally, motifs for CTCF, the mesodermal/PGC-specifying factor Brachyury

[34] and naïve pluripotency factors OCT4 and SOX2 were negatively enriched. (Figure 27C). Motifs positively enriched in the d4 GFP+ PGCLCs relative to EpiLCs also included naïve pluripotency factors SOX2 and OCT4, as well as NANOG, and ESRRB (Figure 27D). Motifs of PGC factors were also positively enriched, including motifs of AP2 $\gamma$  and BLIMP1. Intriguingly, motifs for AP2 $\alpha$  and CTCFL were positively enriched. Thus, analysis of WT motifs which are enriched during the differentiation of EpiLCs towards PGCLC fate demonstrates the negative enrichment of motifs associated with EpiLC TFs and upregulation of motifs associated with PGC factors, as expected.

I then identified DAPs in AP2 $\gamma$ -KO cells relative to WT by cell type, in order to determine if and when during the process of PGC differentiation the KO of AP2 $\gamma$  affects chromatin accessibility. Strikingly, the greatest percentage of DAPs of any of the examined samples was identified in the AP2 $\gamma$ -KO ESCs relative to WT: 8.3% of all peaks were DAPs, with 4.1% (6,463/156,620 peaks) negatively and 4.2% (6,500/156,620 peaks) positively enriched (Figure 28A). This result was surprising, as the AP2 $\gamma$ -KO ESCs demonstrated the least differential gene expression of all examined cell types (see Section 3.2). In comparison, only 0.15% (158/105,876 peaks) of peaks were DAPs in the AP2 $\gamma$ -KO EpiLCs relative to WT, of which 0.11% (111/105,876 peaks) were negatively enriched and 0.044% (47/105,876 peaks) were positively enriched (Figure 28B). Upon differentiation towards PGCLC fate, 0.18% (161/91,003) of the DAPs identified in d4 GFP+ AP2 $\gamma$ -KO cells relative to WT were negatively and 0.077% (70/91,003) were positively enriched (Figure 28C). Although this represents a small increase in the percentage of DAPs relative to what was observed in the AP2 $\gamma$ -KO EpiLCs, it is a surprising result given that AP2 $\gamma$  expression increases in the WT PGCLCs relative to EpiLCs (Figure 12), and also given the fact that AP2 $\gamma$  is a putative pioneer factor. Finally, only seventeen peaks were considered to be DAP in the d6 GFP+ AP2 $\gamma$ -KO cells relative to WT, of which five were negatively and twelve were positively enriched (Figure 28D). The very low number of peaks identified to be DAP may be a result of the dissimilarity of the WT samples previously demonstrated in Figure 25 and Appendices 7.9.1.2 and 7.9.1.4. Overall, it may be concluded that most DAPs may be identified in the AP2 $\gamma$ -KO ESCs, and a very small percentage of peaks are to be considered DAPs as the cells develop further.



**Figure 28: Differential chromatin accessibility is greatest in AP2 $\gamma$ -KO ESCs.** MD plots depicting DAPs identified to be significant by EdgeR (FDR < 0.05, fc not considered; it should be noted that this significance is less stringent than that used for analysis in this work). Red indicates positive enrichment, while blue indicates negative enrichment. Log fold-change (LogFC) and average expression are indicated. MD plots represent: A) DAPs found between AP2 $\gamma$ -KO and WT ESCs, B) DAPs found between AP2 $\gamma$ -KO and WT EpiLCs, C) DAPs found between AP2 $\gamma$ -KO and WT d4 GFP<sup>+</sup> cells, and D) DAPs found between AP2 $\gamma$ -KO and WT d6 GFP<sup>+</sup> cells.

Subsequently, I began the characterization of the AP2 $\gamma$ -KO ESC DAPs via motif enrichment analysis, and additionally annotated the DAPs to the nearest TSS using the HOMER function `annotatePeaks.pl`. This was performed to better understand why such a high percentage of peaks were differentially accessible in the cells which demonstrated the least differential gene expression. Annotation of the peaks

demonstrated that 4,076 genes were annotated to negatively enriched peaks while 4,395 genes were annotated positively enriched peaks (Appendix 7.9.1.7). It should be noted that annotation proceeds based on the nearest TSS relative to the identified peak [221], and thus the annotation may lack functional relevance. Of these genes, only 2.6% (107/4,076) of the genes associated with negatively enriched peaks and 2.9% of the genes associated with positively enriched peaks were differentially expressed in the AP2 $\gamma$ -KO ESC RNA-seq data. It thus appears that the altered accessibility of a peak in the AP2 $\gamma$ -KO ESCs has little overall effect upon gene expression. As many annotated peaks are located in intergenic regions and thus may not have a functional relationship to their annotated gene, I compared DAPs annotated to promoter regions with DE genes identified in RNA-seq. Of 85 promoters identified in the negatively enriched DAPs, only 6 corresponded to DE genes in the RNA-seq data (7.1%), while 294 promoters annotated to positively enriched DAPs corresponded to only 20 (6.8%) genes which were found to be differentially expressed in the RNA-seq data. It should be noted that one gene identified to have a differentially accessible promoter in addition to DE gene expression was *Dppa3*, which had a negatively enriched peak at the promoter and also exhibited downregulation in AP2 $\gamma$ -KO ESCs (Figure 12). This result strongly suggests that AP2 $\gamma$  directly regulates the expression of *Dppa3* by binding to the promoter. Aside from the example of *Dppa3*, it can generally be concluded that the altered accessibility of a promoter in the AP2 $\gamma$ -KO ESCs generally does not correspond to differential expression of the target gene.

Known motif enrichment was then analyzed in the negatively and positively enriched DAPs of the AP2 $\gamma$ -KO ESCs relative to WT, and numerous SOX (SRY-box) family TFs were identified in the top ten negatively enriched AP2 $\gamma$ -KO ESCs motifs, including naïve pluripotency factor SOX2 and endodermal marker SOX17 (SRY-box transcription factor 17; Figure 29A). It is to be noted that the only DE (adj. p. < 0.01, fc +/- 1.5x) SOX gene in the AP2 $\gamma$ -KO ESC RNA-seq data is *Sox21* (SRY-box transcription factor 21), which is downregulated. Motifs belonging to AP2 family members AP2 $\alpha$  and AP2 $\gamma$  represented the two most negatively enriched motifs, while the naïve pluripotency factor NR5A2 was also identified. Numerous TEAD (transcriptional enhanced associate domain) family factors including key TE factor TEAD4 were found in the top ten most positively enriched motifs in the KO ESCs relative to the WT (Figure 29B), as well as motifs for oncogene Myc, 3D chromatin

regulators CTCF and CTCFL, and GATA factors GATA3 and GATA4, which are markers of TE and VE/PE fate, respectively. Broadly, the negative enrichment of the AP2 $\gamma$  motif supports the possibility of AP2 $\gamma$  functioning as a pioneer factor in the WT ESCs, as pioneer factors bind to their own motifs on closed chromatin in order to begin the process of opening chromatin. Further, the enrichment of SOX and TEAD motifs in DAPs is intriguing and suggests that the loss of AP2 $\gamma$  may affect the ability of certain TF families to access and thus regulate target genes.

A Negatively enriched motifs in DE peaks (EdgeR ESC KO-WT)				B Positively enriched motifs in DE peaks (EdgeR ESC KO-WT)			
Rank:	Motif:	Name:	P-Value:	Rank:	Motif:	Name:	P-Value:
1		SOX2	1e-311	1		TEAD	1e-541
2		AP2 $\alpha$	1e-274	2		TEAD1	1e-534
3		AP2 $\gamma$	1e-254	3		TEAD4	1e-519
4		OCT4/SOX2/ TCF/NANOG	1e-250	4		TEAD3	1e-493
5		SOX3	1e-248	5		TEAD2	1e-442
6		SOX15	1e-226	6		Jun-AP1	1e-430
7		SOX17	1e-211	7		FOSL2	1e-402
8		SOX21	1e-189	8		FRA2	1e-366
9		SOX10	1e-173	9		Fos	1e-358
10		LEF1	1e-170	10		FRA1	1e-353
29		NR5A2	1e-75	77		Myc	1e-21
				111		CTCF	1e-10
				114		GATA4	1e-10
				156		CTCF	1e-6
				185		GATA3	1e-3

**Figure 29: AP2 $\gamma$  and AP2 $\alpha$  are negatively enriched in DAPs of AP2 $\gamma$ -KO ESCs.** Motif enrichment analysis of DAPs negatively or positively enriched in AP2 $\gamma$ -KO ESCs relative to WT. The rank and p-value calculated by HOMER are indicated, and significance is determined by p-value < 1e-3. A) Negatively enriched motifs. B) Positively enriched motifs.

Next, I investigated the annotation and motif enrichment of DAPs identified in AP2 $\gamma$ -KO EpiLCs. Although very few DAPs were identified (Figure 28B), I nevertheless wanted to examine if the annotation of DAPs had any relevance to gene expression and if the identified DAPs also demonstrated TF family-specific enrichment. 104 genes were annotated to negatively enriched DAPs, of which only 12 (11.5%) were considered DE, while 47 genes were annotated to positively enriched DAPs, of which only 7 (14.9%) were considered DE (Appendix 7.9.1.7). While this demonstrates

that a greater percentage of annotated genes corresponded to differentially expressed genes in the AP2 $\gamma$ -KO EpiLCs as compared to in the AP2 $\gamma$ -KO ESCs, as the number of overall DAPs is small the percentages may lack functional relevance. Of the 6 negatively enriched and 1 positively enriched DAPs annotated to promoters, none corresponded to DE genes identified in the RNA-seq dataset. Thus, the presence of DAPs in the AP2 $\gamma$ -KO ESCs does not appear to correspond to differential gene expression. Motif enrichment analysis also provided very few results: only two motifs were identified to be significant in the negatively enriched DAPs, belonging to naïve pluripotency factor KLF4 and EKLF (Erythroid Krüppel-like factor), a TF which forms active chromatin hubs (ACHs) during erythropoiesis [259] (Figure 30A). Only nine motifs were found to be significant in the positively enriched DAPs, of which four were associated with POU (Pituitary-specific Pit-1, octamer transcription factor proteins Oct-1 and Oct-2, and Unc-86 transcription factor) class family factors (Figure 30B). The most significantly positively enriched motif belonged to NFY (Nuclear transcription factor Y, also known as NF-Y), a TF enriched in the nucleosome depleted regions of promoters [260]. As very few DAPs were identified, and those that could be were annotated to very few genes and enriched for very few motifs, it may be concluded that the reduced expression AP2 $\gamma$  in the EpiLCs corresponds to few global changes in genome accessibility.

I then aimed to characterize the AP2 $\gamma$ -KO d4 GFP+ DAPs to determine if genomic accessibility was altered, although as previously noted, only a small percentage of DAPs were identified in the AP2 $\gamma$ -KO cells at this stage (Figure 28C). 147 genes were annotated to the negatively enriched DAPs, of which 22 (15.0%) corresponded to differentially expressed genes, while 67 genes were annotated to the positively enriched DAPs, of which 8 (11.9%) corresponded to differentially expressed genes (Appendix 7.9.1.7). Only four promoter regions were annotated to DAPs, and corresponded to negatively enriched DAPs. Three of the four annotated promoters did in fact correspond to genes considered to be DE in the AP2 $\gamma$ -KO d4 GFP+ RNA-seq data, namely *Tfap2c* and the imprinted genes *Peg13* (Paternally expressed 13) and *Mest* (Mesoderm-specific transcript homolog). The IGV profiles demonstrating the DAPs are striking for *Peg13* and *Mest*, as complete depletion of accessible peaks is demonstrated in the AP2 $\gamma$ -KO cells relative to WT in all cell types (Appendix 7.9.1.6A-B). This finding suggests that AP2 $\gamma$  may directly regulate itself as well as some imprinted genes during PGCLC differentiation, although this effect could be

indirect, *i.e.* AP2 $\gamma$  may directly regulate other pioneer factors which in turn directly regulate these genes of interest.

A Negatively enriched motifs in DE peaks (EdgeR EpiLC KO-WT)				B Positively enriched motifs in DE peaks (EdgeR EpiLC KO-WT)			
Rank:	Motif:	Name:	P-Value:	Rank:	Motif:	Name:	P-Value:
1		KLF4	1e-3	1		NFY	1e-4
2		EKLF	1e-3	2		OCT4	1e-3
				3		KLF1	1e-3
				4		NF-E2	1e-3
				5		BRN1	1e-3
				6		OCT:OCT	1e-3
				7		OCT6	1e-3
				8		KLF3	1e-3
				9		DMRT1	1e-3

**Figure 30: POU family motifs are positively enriched in DAPs of AP2 $\gamma$ -KO EpiLCs.** Motif enrichment analysis of DAPs negatively or positively enriched in AP2 $\gamma$ -KO EpiLCs relative to WT. The rank and p-value calculated by HOMER are indicated, and significance is determined by p-value < 1e-3. A) Negatively enriched motifs. B) Positively enriched motifs.

Motif enrichment of AP2 $\gamma$ -KO d4 GFP<sup>+</sup> cells, although performed on a very small number of DAPs, detected motifs with more functional relevance to the cell types studied in this work than those identified in the AP2 $\gamma$ -KO EpiLC DAPs. The two most negatively enriched motifs detected in DAPs were AP2 $\alpha$  and AP2 $\gamma$ . Additionally, numerous KLF (Krüppel-like factor) family motifs were also negatively enriched (Figure 31A). The positively enriched motifs detected in DAPs contained numerous motifs corresponding to SOX family TFs as well as motifs corresponding to the ZIC (Zinc finger of the cerebellum) TF family and TEAD (TEA domain) TF family (Figure 31B). As the SOX TF family has been shown to exert pioneer factor activity, I examined the RNA-seq data for the expression of *Sox* genes. Indeed, several SOX motifs found to be positively enriched in the AP2 $\gamma$ -KO d4 GFP<sup>+</sup> DAPs corresponded to genes which were upregulated in the DE AP2 $\gamma$ -KO d4 GFP<sup>+</sup> RNA-seq data, including *Sox1*, *Sox3*, *Sox4*, *Sox11*, and *Sox13*. Generally, expression of SOX family factors appears to increase with differentiation towards PGC fate in AP2 $\gamma$ -KO cells relative to WT (Appendix 7.9.3.2.) Further, TEAD family genes *Tead1*, *Tead2*, and *Tead3* are not considered significantly DE by d4 in the AP2 $\gamma$ -KO cells, however they are significantly upregulated by d6 in the AP2 $\gamma$ -KO cells (Appendix 7.9.3.3). This result therefore suggest that in the d4 GFP<sup>+</sup> cells, the absence of AP2 $\gamma$  results



in upregulation of SOX genes and corresponding pioneer factor activity, resulting in the increased accessibility of SOX TF family motifs. Additionally, the positive enrichment of TEAD family motifs precedes the upregulation of TEAD factors on d6 in the AP2 $\gamma$ -KO GFP+ cells.



A Negatively enriched motifs in DE peaks (EdgeR d4 GFP+ KO-WT)				B Positively enriched motifs in DE peaks (EdgeR d4 GFP+ KO-WT)			
Rank:	Motif:	Name:	P-Value:	Rank:	Motif:	Name:	P-Value:
1		AP2 $\alpha$	1e-14	1		SOX3	1e-10
2		AP2 $\gamma$	1e-11	2		SOX21	1e-8
3		KLF4	1e-4	3		SOX6	1e-8
4		KLF1	1e-4	4		SOX2	1e-7
5		OCT4/SOX2/ TCF/NANOG	1e-4	5		ZIC	1e-7
6		SP2	1e-4	6		SOX10	1e-6
7		EKLF	1e-4	7		SOX4	1e-6
8		KLF5	1e-4	8		TEAD	1e-6
9		KLF6	1e-3	9		ZIC2	1e-5
10		KLF3	1e-3	10		SOX15	1e-5
				11		TEAD4	1e-5
				12		SOX17	1e-5
				24		OTX2	1e-3

**Figure 31: AP2 $\gamma$  and AP2 $\alpha$  are negatively enriched in DAPs of AP2 $\gamma$ -KO d4 GFP+ cells.** Motif enrichment analysis of DAPs negatively or positively enriched in AP2 $\gamma$ -KO d4 GFP+ cells relative to WT. The rank and p-value calculated by HOMER are indicated, and significance is determined by p-value < 1e-3. A) Negatively enriched motifs. B) Positively enriched motifs.

Finally, as only 17 peaks were considered to be DAP in the AP2 $\gamma$ -KO d6 GFP+ cells (Figure 28D), it was difficult to characterize the state of chromatin accessibility in these cells. Genes annotated to these DAPs did not correspond to DE genes in the AP2 $\gamma$ -KO d6 GFP+ RNA-seq data. No significant motifs could be identified for the negatively enriched DAPs, while only two could be identified for the positively enriched DAPs, namely p73 (Transformation related protein 73, also known as TRP63), an oncogenic transcription factor associated with cancer stem cells [261] and SOX9 (Figure 32). As the WT d6 GFP+ samples were so dissimilar, it is difficult to assess the relevance of this data. It cannot be concluded that so few DAPs or enriched motifs resulted from the unaltered chromatin accessibility of these cells, rather it is more likely that the dissimilarity of the WT samples do not allow for the correct identification of significant DAPs relative to AP2 $\gamma$ -KO d6 GFP+ cells.



Positively enriched motifs in DE peaks  
(EdgeR d6 GFP+ KO-WT)

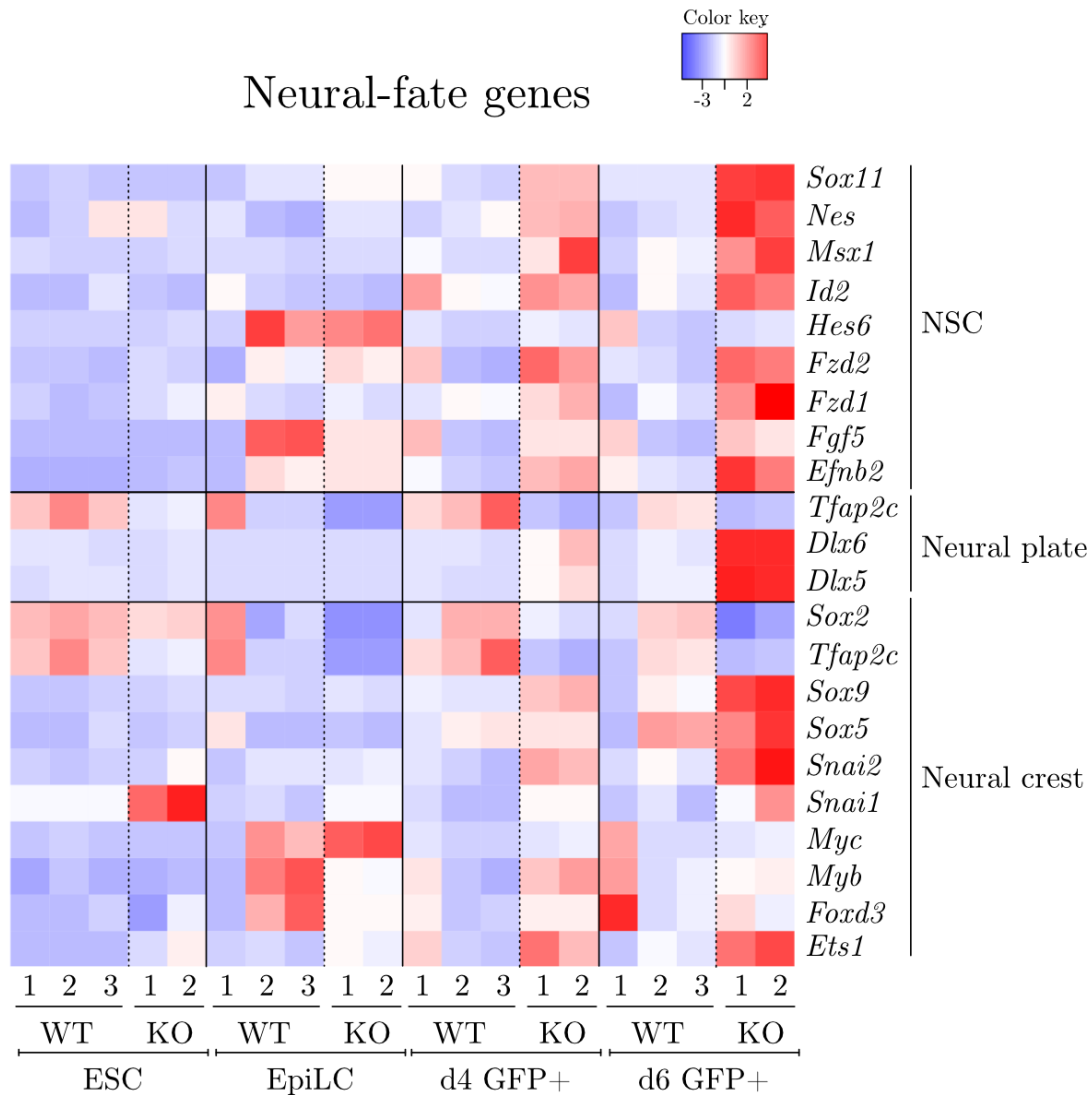
Rank:	Motif:	Name:	P-Value:
1		p73	1e-3
2		SOX9	1e-3

**Figure 32: Only SOX9 and p73 are positively enriched in DAPs of AP2 $\gamma$ -KO d6 GFP+ cells.** Motif enrichment analysis of positively enriched DAPs in AP2 $\gamma$ -KO d6 GFP+ cells relative to WT. The rank and p-value calculated by HOMER are indicated, and significance is determined by p-value < 1e-3.

Overall, ATAC-seq analysis has demonstrated that DNA binding motifs for AP2 $\gamma$  as well as closely-related AP2 family member AP2 $\alpha$  are enriched in naïve- and PGC-specific ATAC peaks in WT cells. Additionally, the motifs of both factors are negatively enriched in AP2 $\gamma$ -KO ESCs and d4 GFP+ cells. As WT ESCs and d4 GFP+ cells upregulate *Tfap2c*, this result therefore suggests that AP2 $\gamma$  may indeed be behaving as a pioneer factor in WT ESCs and d4 GFP+ by opening chromatin containing the AP2 $\gamma$  or AP2 $\alpha$  binding motifs, and in its absence these regions demonstrate reduced accessibility. The negative enrichment of AP2 $\gamma$ /AP2 $\alpha$  binding motifs is not observed in AP2 $\gamma$ -KO EpiLCs, and as AP2 $\gamma$  is downregulated in EpiLCs, this result indicates that when AP2 $\gamma$  is not strongly expressed, there is not difference in accessibility of chromatin bearing its motif. Further, several classes of TF families, including SOX, ZIC, KLF, and TEAD families, were strongly enriched in the AP2 $\gamma$ -KO ESC and d4 GFP+ DAPs. This result raises the possibility that AP2 $\gamma$  may directly or indirectly regulate the accessibility of chromatin for certain classes of TF in cell-type specific contexts. Finally, it is of note that most DAPs were identified in the AP2 $\gamma$ -KO ESCs and a very small number were identified in the AP2 $\gamma$ -KO d4 GFP+ cells, as more genes are differentially expressed in the AP2 $\gamma$ -KO d4 GFP+ cells relative to the AP2 $\gamma$ -KO ESCs. The paradoxical nature of this result raises the possibility that a redundant pioneer factor is upregulated in the AP2 $\gamma$ -KO d4 GFP+ cells which is able to compensate for the pioneer factor activity of AP2 $\gamma$ , yet cannot compensate for its gene regulatory activities.

### 3.3.3 Expression of markers of neural fate in AP2 $\gamma$ -KO cells

The intriguing finding that AP2 $\gamma$ -KO d4 GFP+ cells upregulate SOX family TF factors and demonstrate positive enrichment for SOX, TEAD, and ZIC family TF binding sites corresponds to a finding by Pastor *et al.*: primed human AP2 $\gamma$ -KO ESCs fail to revert to naïve pluripotency, and rather exhibit positive enrichment



**Figure 33: Expression of neural fate genes increases with development towards PG-CLC fate following KO of AP2 $\gamma$ .** Heatmap of genes associated with neural stem cells (NSCs), neural plate development, or neural crest fate. Red indicates increased expression, blue indicates downregulation. Significance (adj. p. < 0.01, fc +/- 1.5x) determined in AP2 $\gamma$ -KO d6 GFP+ cells against WT. Sample and gene names are indicated.

of SOX and ZIC family TF binding motifs, which they conclude indicates a skew towards neural fate [257]. As it has additionally been shown that the KD of AP2 $\gamma$  primes mouse ESCs towards neural fate upon EB differentiation [262], I examined the expression of neural crest, neural plate, and neural stem cell (NSC) markers in the RNA-seq dataset to determine if the upregulation of SOX and ZIC factors corresponded to the upregulation of neural markers. Indeed, numerous markers of the neural crest, plate, and NSCs were found to be significantly upregulated in the AP2 $\gamma$ -KO d4/d6 GFP+ cells

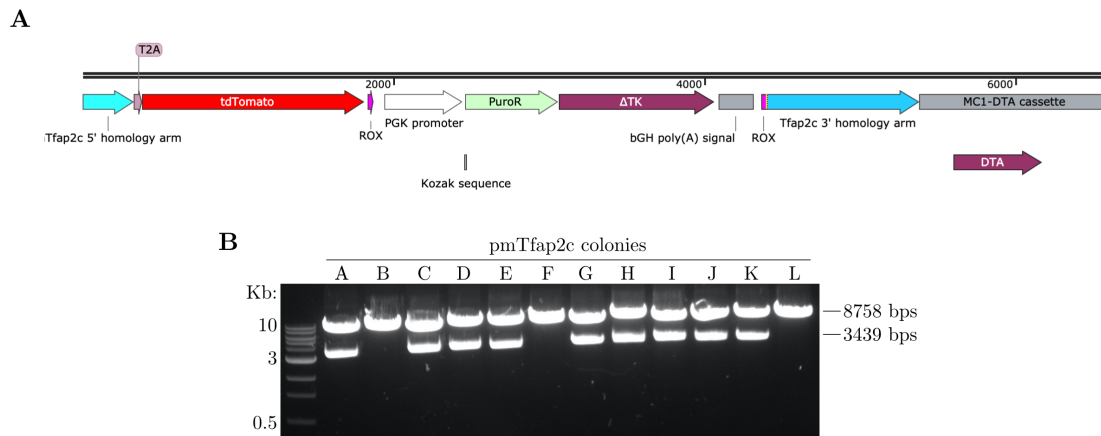
(Figure 58). This finding underlines the fact that AP2 $\gamma$ -KO d4/d6 GFP<sup>+</sup> cells fail to maintain the PGCLC transcriptional program, and instead aberrantly express markers of somatic fate. Further, it raises the question of whether the downregulation of PGC and naïve pluripotency markers and corresponding upregulation of SOX family factors and neural markers in AP2 $\gamma$ -KO d4/d6 GFP<sup>+</sup> cells may also be explained by the Pastor *et al.* finding, that AP2 $\gamma$  behaves as a pioneer factor and is required to open naïve-specific peaks necessary for the re-expression of naïve pluripotency factors.

## 3.4 *Tfap2c* Tiling screen

### 3.4.1 Development of AP2 $\gamma$ reporter cell line

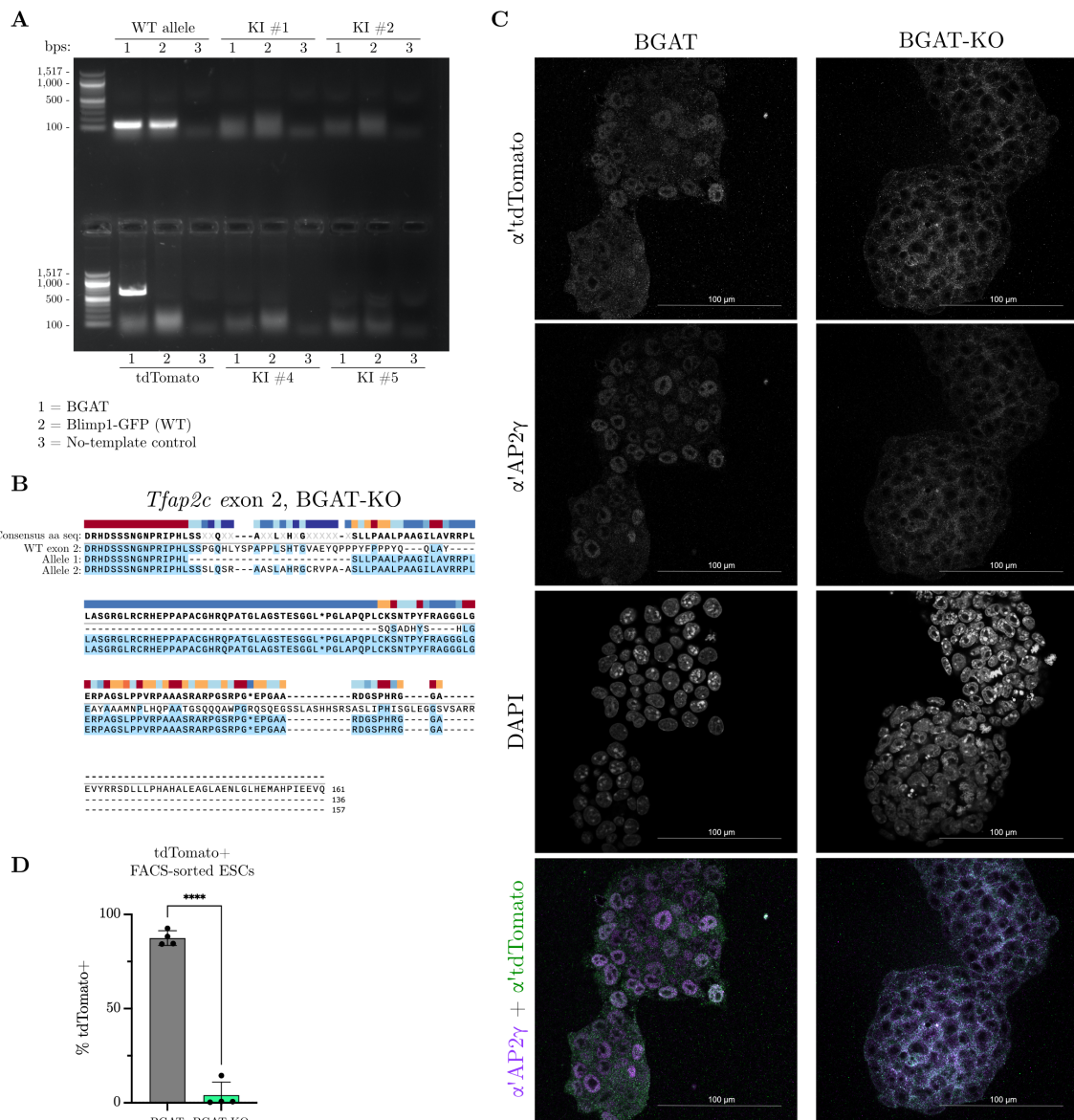
The expression of *Tfap2c* corresponds to a “V” or “high-low-high” expression profile, meaning that *Tfap2c* is expressed in ESCs, downregulated in EpiLCs, and upregulated in PGCLCs (Figure 12). I thus designed a *Tfap2c* reporter cell line with two aims in mind: 1) to study *Tfap2c* transcriptional dynamics via cell type-specific fluorescence reporting, and 2) to utilize the cell line in the *Tfap2c* tiling screen to report the downregulation or loss of *Tfap2c* expression. Towards this end, I employed CRISPR/Cas9 to introduce a double-strand break (DSB) into the seventh and final exon of *Tfap2c* immediately upstream of the stop codon, in order to introduce the tdTomato fluorescent reporter gene using a homology directed repair (HDR) repair template. The repair template, plasmid pmTfap2c (pmTfap2c-T2A-tdTomato-RoxPGKPuro; Appendix 7.8.6), was designed using the phNANOS3 plasmid [175] as a design template, with the fluorescence reporter cassette and selection cassette being entirely copied from the phNANOS3 plasmid (Appendix 7.8.2). A linear sequence from plasmid pmTfap2c is shown in Figure 34A and demonstrates the planned design of the plasmid. The sequences of the 5' and 3' homology arms flanking the stop codon of *Tfap2c*, as well as the *Tdtomato* red fluorescent reporter cassette are indicated. Between the 5' homology arm and the *Tdtomato* cassette, a *Thosea asigna* virus 2A (T2A) sequence was inserted, which allows for the internal processing of the protein tdTomato from the expressed AP2 $\gamma$  protein [263], thus alleviating the potential risk of the reporter actively changing the conformation of the tagged protein. A roxP-flanked selection cassette containing a mouse phosphoglycerate kinase 1 (PGK) promoter, a puromycin resistance cassette, a  $\Delta$ TK cassette, and a bovine growth hormone polyadenylation (bGH poly(A)) signal was located downstream of the *Tdtomato* cassette. The puromycin resistance gene confers resistance to the eukaryotic antibiotic puromycin, and thus allows for the selection of correctly transfected cells which contained the knock-in sequence. ROX sites are recombination target sites of Dre recombinase [183] - the ROX sites flanking the selection cassette allow for the possibility of the targeted excision of the selection cassette following completed puromycin selection.  $\Delta$ TK is a truncated Herpes simplex virus type 1

thymidine kinase gene which converts Ganciclovir into a cytotoxic chemical, thus acting as a “suicide gene” and killing the cell [264].  $\Delta$ TK can thus be used as a control to determine if the Dre recombination of the selection cassette was successful. A diphtheria toxin A (DTA) gene was inserted under the control of an MC1 promoter to negatively select cells which aberrantly genomically inserted plasmid sequences located outside of the homology arms.



**Figure 34: Design and identification of plasmid pmTfap2c.** A) Plasmid pmTfap2c is comprised of the following features: the 5' *Tfap2c* homology arm (turquoise), a tdTomato cassette (red) separated from the *Tfap2c* 5' homology arm by a T2A site (pink), a PGK promoter (white) upstream of a puromycin resistance gene (green), a  $\Delta$ TK gene (purple) immediately upstream of a bGH pol(A) signal (grey), the 3' *Tfap2c* homology arm (blue) separated from the poly(A), and an MC1-DTA cassette (grey). The selection cassette is flanked by two ROX sites (pink). B) pmTfap2c was Gibson-cloned and transformed into DH5 $\alpha$  transformation-competent bacteria. Single clones were isolated and resulting plasmids were purified and digested with SnaBI/XhoI to identify correct clones. Here, two bands indicate correct cloned pmTfap2c plasmids, with sized in kb/bps indicated.

I cloned the plasmid using the Gibson cloning method [181], and RE digested cloned plasmids isolated from bacterial colonies with SnaBI/XhoI to identify correct ligation of fragments as well as correct plasmid size (Figure 34B). One plasmid clone was completely characterized and found to have all fragments correctly inserted (plasmid A, Figure 34B). I then used Sanger sequencing to validate that the cloned sequences were directionally and sequentially correct.



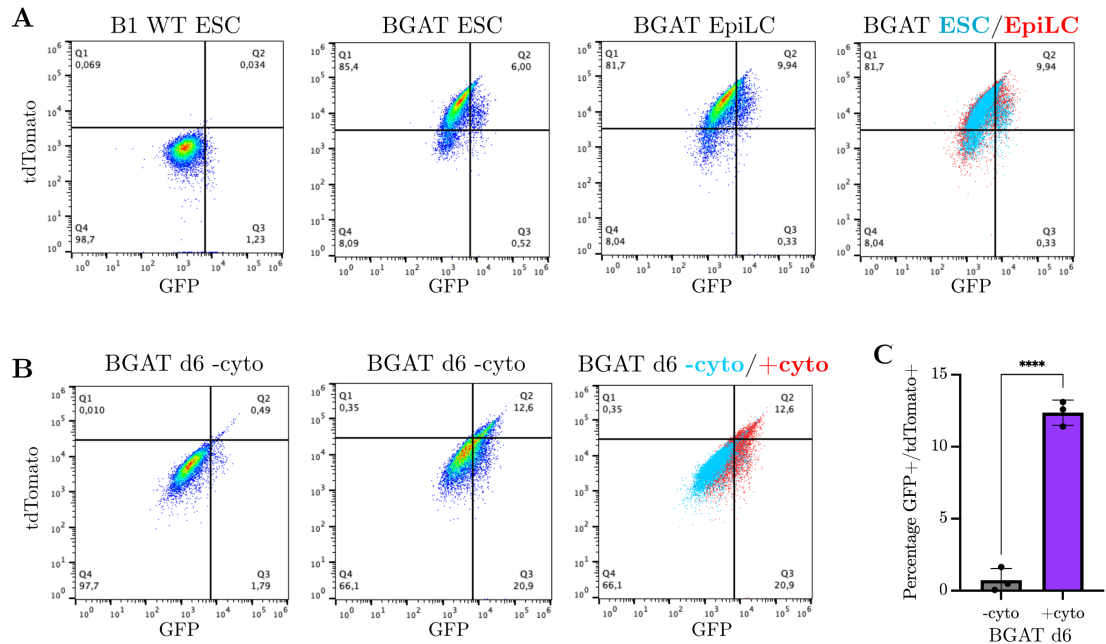
**Figure 35: The BGAT reporter line demonstrates reduced tdTomato expression following AP2 $\gamma$ -KO.** A) gDNA isolated from B1 and BGAT cells was genotyped for the presence of the WT *Tfap2c* allele, the tdTomato cassette, and four combinations of knock-in (KI) primers. No-template control was included, and sized in bps are indicated. B) pMiniT™ cloning and Sanger sequencing was translated to aa sequences in SnapGene®. Premature stop codons are indicated as asterisks (\*), and blue-highlighting demonstrates consensus between the WT and AP2 $\gamma$ -KO aa sequences. C) 40x confocal images depicting the ICC staining of BGAT and BGAT-KO ESCs.  $\alpha'$ AP2 $\gamma$ -Alexa-488 is shown in green,  $\alpha'$ tdTomato-Alexa-647 is shown in purple, and DAPI is shown in blue, with scale bar of 100  $\mu$ m indicated. D) The percentage of tdTomato+ BGAT and BGAT-KO ESCs, potentially reporting AP2 $\gamma$  expression, is indicated. N = 3 for all samples. Significance (p < 0.05) is indicated, and was determined using the unpaired t-test in Prism 9.

Next, I transfected B1 WT cells with the repair template pmTfap2c as well as the plasmid pX330-a'mTfap2c-stop-g5, containing the Cas9 gene as well as a gRNA targeting Cas9

to the region immediately upstream of the *Tfap2c* stop codon. Following two days of selection in puromycin, I sorted the transfected cells as single cells using FACS. Lines derived from single clones were characterized based on live cell tdTomato+ fluorescence. I further transfected tdTomato+ clonal lines with pCAGGS-Dre-IH, pPY-CAG-PBase, and pPB-CAG-Cas9-IRES-Hygro to both excise the ROX-flanked selection cassette and genomically insert Cas9 via PiggyBac transposase-directed activity. I then sorted the resulting cell lines again as single cells using FACS. The individual clonal cell line characterized in this work was termed Blimp1-meGFP AP2 $\gamma$ -tdTomato gRNA #5 clone 2A ESC line (BGAT). I initially attempted to characterize the BGAT line via genotyping PCR, however only the WT allele and the *Tdtomato* insertion could be detected, and PCR reactions targeting the knock-in region for amplification, wherein one primer bound to the WT genomic DNA and the other primer bound to the inserted DNA, routinely failed (Figure 35A).

As an alternative means to genotype the cells, I employed CRISPR/Cas9 to knock-out *Tfap2c* and observe the subsequent effect upon tdTomato expression. To this end, I transfected BGAT cells with plasmids pX330-a'mTfap2c-exon2-g2 as previously described (Section 3.1) and sorted single cells using FACS to generate individual clonal lines. Sanger sequencing of pMiniT<sup>TM</sup> plasmids inserted with the gRNA-targeted *Tfap2c* exon two genomic sequence revealed that one KO line was an AP2 $\gamma$ -KO line, with an allele containing a 74 bp deletion and another allele containing an 11 bp deletion (Figure 35B). Both mutations generate frameshifts which result in predicted early stop codons. This line was henceforth termed Blimp1-meGFP AP2 $\gamma$ -tdTomato gRNA #5 clone A AP2 $\gamma$ -KO clone #14 ESC line (BGAT-KO). I then stained the BGAT and BGAT-KO cells with tdTomato- and AP2 $\gamma$ -targeting antibodies and imaged the IF samples at the confocal microscope, which demonstrated that in the BGAT line tdTomato expression colocalizes with AP2 $\gamma$  expression, while in the BGAT-KO line the nuclear AP2 $\gamma$  stain is completely ablated and the tdTomato stain is completely lost (Figure 35C). It should be noted, however, that the tdTomato stain observed in the BGAT cells is nuclear, although the tdTomato protein should be processed from the AP2 $\gamma$  protein via the T2A site, and has no nuclear localization tag. To quantify the loss in tdTomato fluorescence, I sorted the BGAT and BGAT-KO ESCs using FACS. On average, 87.4% of the BGAT cells were tdTomato+, while only 4.02% of the BGAT-KO cells were tdTomato+, a significant ( $p < 0.05$ ) reduction (Figure 35D).

I next characterized the BGAT cells as EpiLCs and d6 PGCLCs. BGAT ESC and EpiLCs were sorted using FACS and directly compared; BGAT EpiLCs did not lose red fluorescence compared to BGAT ESCs (Figure 36A), which was inconsistent with RNA-seq data demonstrating that *Tfap2c* expression is reduced in EpiLCs relative to ESCs (Figure 12). However, on d6 of PGCLC differentiation, a population of tdTomato+/GFP+ cells emerged



**Figure 36: Expression of tdTomato in BGAT cells during PGCLC differentiation.** A) FACS profiles of B1 WT ESCs, BGAT ESCs, and BGAT EpiLCs with GFP and tdTomato fluorescence indicated. Gating was performed around B1 WT ESCs to indicate tdTomato<sup>+</sup>/GFP<sup>+</sup> fluorescence in the daughter BGAT line. The overlay of BGAT ESC and EpiLC profiles is included, with the colored population indicating cell type - ESC in blue, EpiLC in red. B) FACS profiles of BGAT d6 -cyto and +cyto cell populations with GFP and tdTomato fluorescence indicated. Gating was performed around the d6 BGAT -cyto population to identify gained GFP<sup>+</sup>/tdTomato<sup>+</sup> fluorescence upon the addition of cytokines, and the overlay of +/-cyto populations indicates -cyto cells in blue and +cyto cells in red. C) Statistical analysis of the percentage of BGAT d6 tdTomato<sup>+</sup>/GFP<sup>+</sup> cells +/-cyto, with significance ( $p < 0.05$ ) calculated via unpaired t-test in Prism 9 indicated.  $N = 3$  for all samples.

in the +cyto EBs which was not observed in the -cyto EBs (Figure 36B). On average, this resulted in 12.4% tdTomato<sup>+</sup>/GFP<sup>+</sup> BGAT +cyto d6 cells, compared to 0.73% double-positive cells in the -cyto negative control condition (Figure 36C).

In consideration of the fact that the knock-in allele could not be genotyped, the tdTomato was localized aberrantly to the nucleus, and no change in tdTomato expression was seen between ESCs and EpiLCs, I determined that the BGAT cell line could not be considered a reliable reporter of AP2 $\gamma$ , and thus decided to find an alternative cell line in which to perform the screen. A suitable alternative was presented by the SGET cells, which have previously been shown to demonstrate reduced GFP expression in the S-KO line (Figure 22). The reduction in GFP<sup>+</sup> cells therefore indirectly reports the loss of AP2 $\gamma$  expression. I chose this line for use in the *Tfap2c* tiling screen, and subsequently transfected SGET ESCs with pPY-CAG-PBase and pPB-CAG-Cas9-IRES-Hygro to genomically insert the

Cas9 gene via the action of the PiggyBac transposase. The resulting cell line was named Stella-eGFP ESG1-tdTomato Cas9 clone # 10 (SGET-Cas9) ESC line.

### 3.4.2 *Tfap2c* tiling screen design and preparation

The *Tfap2c* CRISPR/Cas9 tiling screen required a library of gRNAs to introduce mutations into the *Tfap2c* topologically active domain (TAD) and other targeted regions of interest to potentially knockout, and thus allow the for the identification of, *cis*-regulatory elements (CREs). The generation of gRNAs for the tiling library was designed by Xiaojuan Li (of the Günesdogan Lab) using the Joung *et al.* script [156; 224]. This design included the TAD where the gene *Tfap2c* is located, TAD 325 (chr2:172,454,500 - 172,814,499), which was originally identified using Hi-C data [265]. The TAD of *Tfap2c* was chosen as the primary target for gRNA library tiling, as it is expected that most interactions between a CRE and the promoter of a gene occur within the same TAD [160]. The nearest quarter of each TAD neighboring TAD 325 were included in the gRNA design process as negative control regions (chr2:171,904,500 - 172,454,500, chr2:172,814,499 - 173,624,499). Unpublished Capture Hi-C data from our lab and the lab of Dr. Stefan Schönfelder at the Babraham Institute [266] has identified twenty-two potential CREs interacting with the promoter of *Tfap2c* in ESCs, and ten potential CREs interacting with the promoter of *Tfap2c* in EpiLCs. The majority of these interactions are located within TAD 325, however seven regions outside of this TAD were identified (chr2:170,246,202 - 170,503,872, chr2:158,575,739 - 158,597,087, chr2:178,627,369 - 178,633,710, chr1:100,238,852 - 100,242,467, chr11:78,007,245 - 78,013,145). These regions were also included during the gRNA design process.

As a control for enhancer detection within the screen, gRNAs targeting eight putative enhancers identified in ESCs [267] were included. Four of the enhancers were functionally validated and demonstrated to be required for maintenance of the ESC fate and were included in the screen as positive controls (chr5:119,656,454 - 119,661,324, chr8:43,351,590 - 43,356,510, chr8:72,290,869 - 72,293,289, chr14:99,323,583 - 99,327,543), while the other four enhancers were not found to be required for ESC fate maintenance and were included in the screen as negative controls (chr8:43,327,503 - 43,332,433, chr14:99,333,716 - 99,341,806, chr17:47,950,985 - 47,959,425, chr18:81,196,705 - 81,198,815). gRNAs targeting key PGC genes *Tfap2c*, *Prdm1*, and *Prdm14*, genes *Dppa3* and *Dppa5a* with fluorescent reporter activity in the Stella-eGFP Esg1-tdTomato (SGET) cell-line used during the screen, and the genes of the fluorescent reporters themselves, *Egfp* and *Tdtomato*, were included in the generation of gRNAs in the library as positive controls. Finally, scrambled gRNAs which do not have a target in the mouse genome were as negative controls, and two five kb non-coding genetic regions with no known relationship to *Tfap2c* expression or regulation were



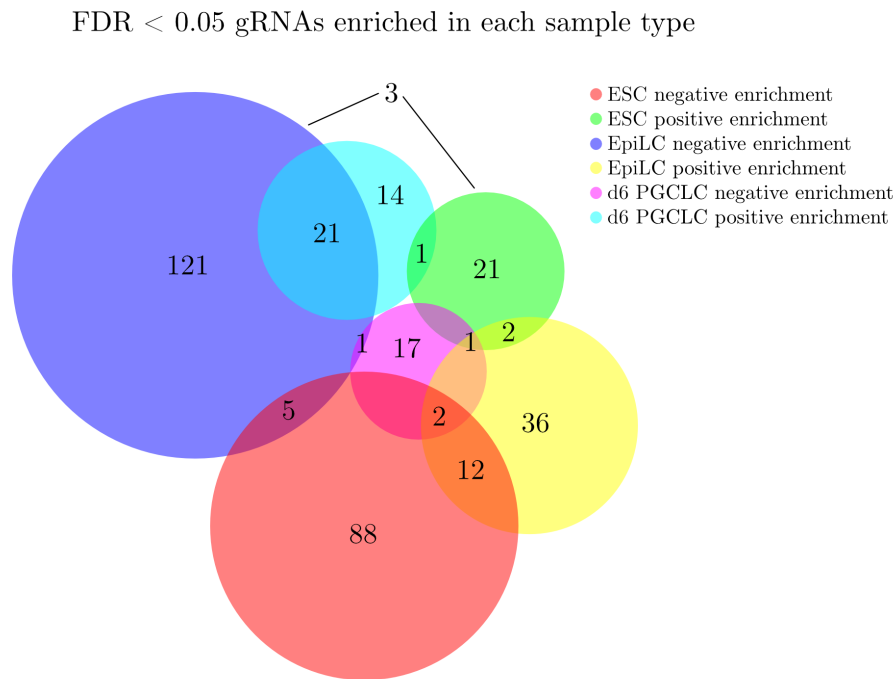
included as regions for negative control gRNA generation (chr9:29,055,657 - 29,060,656; chr13:66,700,268 - 66,705,267).

The gRNA library was ordered as an oligo library, and I cloned it into the plasmid pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W, transformed the resulting plasmid library into electro-competent bacteria, and purified the resulting plasmid library. I then PCR-amplified the gRNA-containing region of the plasmid and sent the purified amplicon for sequencing on an Illumina Hi-Seq, which produced  $\sim 18$  Mio 150 bp single-end fragments. Sequencing analysis demonstrated that 30,313/30,714 gRNAs were accounted for, in total 98.7% of all gRNAs. 1.3% of gRNAs were lost, which is close to the suggested MaGECK quality control standard of 1% [241]. The Gini index, a value indicating the evenness of distribution, was 0.08335. A value below 0.1 is recommended for the plasmid library control [241]. Thus, the library was considered suitable for further use. It should be noted that one gRNA targeting *Tfap2c* was overrepresented (gRNA chr2\_11723). As individual gRNAs contribute to single viruses which generate single-KO lines in the pooled screen, the overrepresentation of a single gRNA needs to be taken into consideration when analysing screening results.

To prepare lentivirus, I transfected HEK cells with plasmids psPAX2, pMD2.G, and the pKLV2-W-based gRNA library. I then collected and filtered the lentiviral supernatant was, and determined the lentiviral titer by transducing SGET-Cas9 with 0-160  $\mu$ l of lentiviral supernatant per well for twenty-four hours and selecting the cells in puromycin-containing medium for four days. Following selection, I counted the cells, and the proportion of surviving cells to untransduced/puromycin-untreated cells was used to determine the multiplicity of infection (MOI). Two parameters were suboptimal: First, a higher concentration of puromycin was employed than is generally used to transfect ESCs (see Section 2.9.3), which may increase the likelihood that cells transfected with multiple gRNAs were selected [156]. However, the volumes of lentiviral supernatant per ml necessary to calculate the MOI were well within the range suggested by Joung *et al.* [156], indicating that the concentration of puromycin did not affect the MOI determination. Second, the MOI used during the screen to transduce the SGET-Cas9 ESCs was ultimately determined to be 0.3815, slightly higher than the standard value of 0.3. With an MOI of 0.3, 74.1% of cells are untransduced with virus, 22.2% are singly-transduced, *i.e.* one viral particle is predicted to infect a cell, and 3.7% of cells are multiply-transduced [239]. With an MOI of 0.3815, the likelihood of transfecting one cell with multiple viral particles instead of one is slightly increased (1.9%). However, this small difference in MOI is unlikely to affect the overall performance or read-out of the screen.

### 3.4.3 Tiling screen of *Tfap2c* TAD in SGET-Cas9 cells

SGET-Cas9 ESCs were transduced with lentiviral medium for 24 hours, and selected for twelve days in puromycin-containing medium. The cells were subsequently differentiated towards EpiLC fate, and a pellet containing more than  $1.54 \times 10^7$  ESCs was collected. This value was chosen to ensure that at least 500 cells per gRNA were represented in the screen. Following 42 hours of EpiLC differentiation, the cells were induced towards PGCLC fate, and a pellet of at least  $1.54 \times 10^7$  EpiLCs were again collected. After six days of differentiation, cells were FACS-sorted to isolate GFP<sup>+</sup>/tdTomato<sup>-</sup> cells. The transduction and differentiation process proceeded in total three times, and resulted in the collection of  $2.31 \times 10^6$ ,  $4.62 \times 10^6$ , and  $4.84 \times 10^6$  d6 SGET-Cas9 GFP<sup>+</sup>/tdTomato-PGCLCs, resulting in a gRNA coverage of 75, 150, and 159 cells per gRNA, respectively, and resulting on average  $\sim 128$  cells/gRNA.



**Figure 37: Unique and shared gRNAs per cell type and enrichment status.** Venn diagram depicting negatively and positively enriched gRNAs identified to be significant (FDR < 0.05) per cell type analyzed in screen. Overlapping regions without numbers do not indicate shared gRNAs, rather are a spatial necessity of the graph. Cell type and enrichment status are indicated in key. Venn diagram generated using DeepVenn.

To amplify the genomically-inserted gRNA sequences, I collected genomic DNA from each sample pellet for template in the PCR to generate sequencing libraries. The resulting libraries were sequenced, resulting in between  $\sim 10$ -27 Mio single-end reads. All libraries

mapped >78% with the reference gRNAs, and statistics including the number of gRNAs not identified per library and Gini indexes may be found in Appendix 7.9.2.1. I analyzed the sequenced libraries with MaGECK-VISPR MLE (Maximum-likelihood estimation), which identified 107 negatively enriched and 28 positively enriched gRNAs in the ESC samples, 151 negatively enriched and 53 positively enriched gRNAs in the EpiLC samples, and 21 negatively enriched and 36 positively enriched gRNAs in the d6 GFP+/tdTomato- PGCLCS (Figure 37). Multiple gRNAs were enriched between cell types and even between enrichment states, *i.e.* 21 gRNAs which were negatively enriched in EpiLCs were positively enriched in PGCLCS. A visual overview of the *Tfap2c* TAD and the quarters of surrounding TADs may be seen in Figure 38, with the gRNA library coverage and significant gRNAs indicated, and WT ATAC-seq and RNA-seq Bigwig files included for reference. The *Tfap2c* gene is indicated, to demonstrate the coverage of the gRNA library as well as the significant gRNAs targeting the gene.

I then examined the enrichment of control GeCKO gRNAs targeting PGC genes *Tfap2c*, *Prdm14*, *Dppa3*, and *Blimp1*, cell line-specific gene *Dppa5a*, or reporter genes *Gfp* and *Tdtomato* to identify if the controls were appropriately positively- or negatively-enriched in the PGCLCS (positive or negative enrichment is indicated by positive or negative beta score; Table 2). Indeed, in the d6 PGCLCS gRNAs targeting *Dppa3*, *Prdm14*, and *Gfp* were found to be negatively enriched, while *Tdtomato*-targeting gRNAs were found to be positively enriched, all as expected. However, gRNAs targeting *Tfap2c*, *Dppa5a*, and *Prdm1* were not determined to be significant in the d6 PGCLCS. It should be noted that the 5/6 gRNAs targeting *Dppa5a*, although taken from the published GeCKO library, show multiple 100% off-target hits in the genome when aligned with the mm10 mouse genome using the UCSC BLAT (BLAST-like alignment tool) tool (data not shown). This may explain why they were not identified to be positively enriched in the d6 PGCLCS, as was expected. Thus, significant enrichment of most PGC control gRNAs demonstrates that the screen successfully detected most genes detecting PGC fate, and increases the overall confidence that other gRNAs detected in the screen are correctly reporting positive or negative enrichment. However, the failure to detect *Tfap2c*-targeting gRNAs raises questions about the ability of the screen to detect gRNA-targeted CREs of *Tfap2c*.

Although the GeCKO control gRNAs were included to ensure that gRNAs targeting PGC genes could be detected in the d6 PGCLC samples, it is intriguing that gRNAs targeting both *Dppa3* and *Dppa5a* were depleted in ESCs, while *Prdm14*-targeting gRNAs exhibited positive enrichment. The negative enrichment of *Dppa5a* gRNAs may lack functional relevance due to the fact that 5/6 gRNAs map to multiple 100% off-target sites in the genome. Negative enrichment of *Dppa3*-targeting gRNAs was also detected in the EpiLCs,

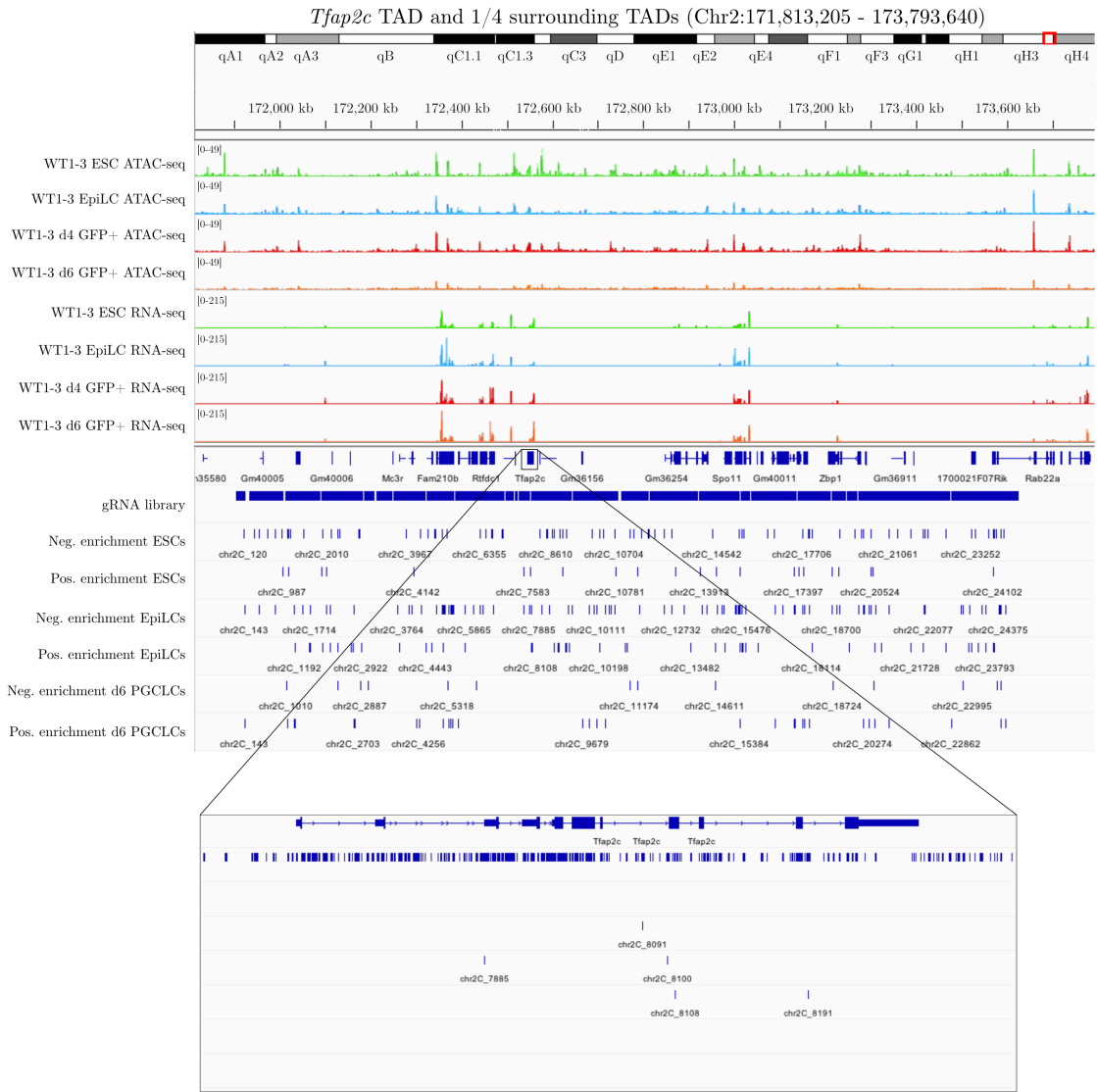
as was the positive enrichment of *Prdm14*-targeting gRNAs. As these gRNAs were not included as controls for ESC or EpiLC fate, the functional relevance of the enrichment they demonstrate cannot easily be determined.

**Table 2:** PGCLC control gRNAs found to be significant in sequenced cell types

Cell type:	Enrichment:	Gene name:	Beta score:	FDR:
ESC	Negative	<i>Dppa3</i>	-0.76	0
		<i>Dppa5a</i>	-0.4	0.022
ESC	Positive	<i>Prdm14</i>	1.9	0
EpiLC	Negative	<i>Dppa3</i>	-0.38	0.0098
EpiLC	Positive	<i>Prdm14</i>	0.36	0.015
d6 PGCLC	Negative	<i>Dppa3</i>	-0.75	0
		<i>Prdm14</i>	-0.71	0
		<i>Gfp</i>	-0.53	0.017
d6 PGCLC	Positive	<i>Tdtomato</i>	0.73	0.017

Once I had determined that most relevant controls in the d6 PGCLCs had exhibited correct enrichment following analysis, I annotated significant gRNAs to genomic features. Interestingly, analysis demonstrates that most gRNAs target intergenic regions (“No features”) and introns, locations where CREs are generally located [135]. I identified hits in ENCODE candidate CREs (cCREs), exons, introns, previously discussed control genes, and a pseudogene (*Ppp4r1l-ps*; Protein phosphatase 4, regulatory subunit 1-like, pseudogene) It was also observed that in every analyzed condition except the positively enriched EpiLC gRNAs, gRNA sequences fall within the coordinates of cCREs. Of note, most exon-targeting gRNAs are negatively enriched in EpiLCs. 25 *Cstf1*-, 16 *Aurka*-, and 10 *Rae1*-targeting gRNAs were negatively enriched in EpiLCs. Enriched gRNAs targeting either *Aurka* or *Cstf1* are visualized in Figure 40. All three genes encode proteins which provide house-keeping functions: *Aurka* (Aurora kinase A) is involved in mitotic spindle regulation and entry into mitosis [268], *Cstf1* (Cleavage stimulation factor, 3' pre-RNA, subunit 1; also known as CstF-50) is a component of a complex which is involved in mRNA poly(A) tail processing [269], and *Rae1* is a mitotic checkpoint regulator required for the correct segregation of chromosomes [270]. The significant negative depletion of gRNAs targeting these genes indicates that they may be required for EpiLC induction, differentiation, or survival.

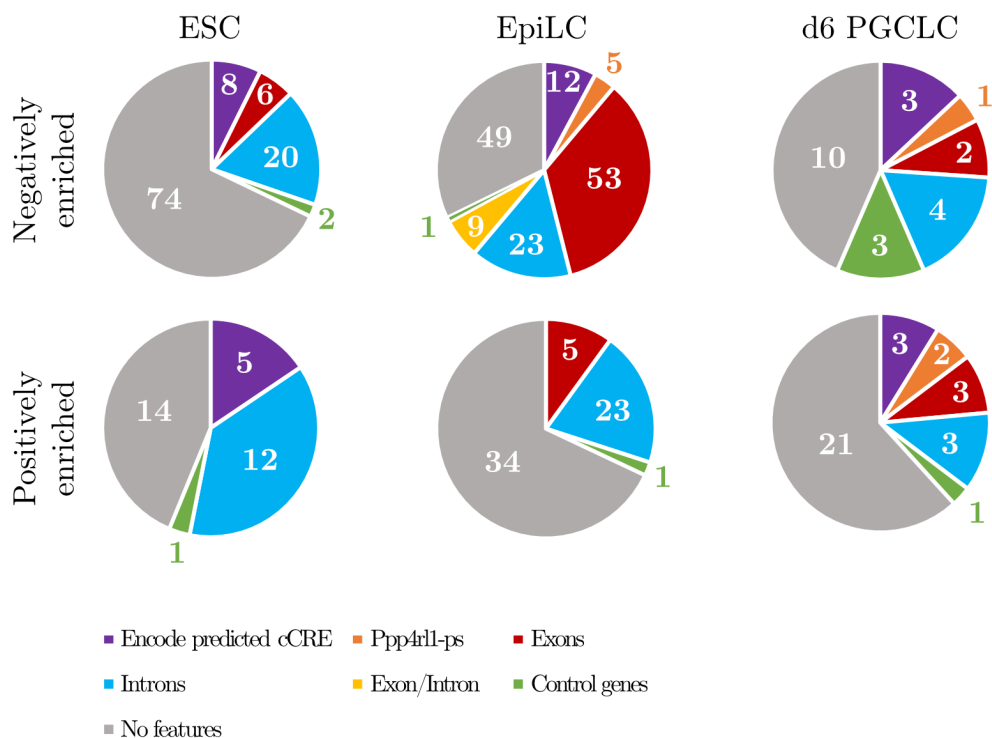
In total, only five gRNAs targeting *Tfap2c* were considered significant (Figure 38), and only one of these gRNAs was exon-targeting (chr2C\_8108). In consideration of the fact that the GeCKO *Tfap2c*-targeting control gRNAs were determined to be not significant in



**Figure 38:** *Tfad2c* TAD and surrounding regions demonstrate coverage of gRNA library and significant hits by cell type and enrichment status. gRNA coverage of the *Tfad2c* TAD is indicated in the lane “gRNA library”. Significantly enriched (FDR < 0.05) gRNAs are identified in each indicated lane by cell type and positive or negative enrichment status. WT ATAC-seq and RNA-seq data is included for reference, with NGS library and cell type indicated. The *Tfad2c* gene is indicated and digitally enlarged to demonstrate significant gRNAs. Image generated in IGV with genes and chromosomal coordinates and increments indicated.

the PGCLC samples, it can be concluded that this screen did not detect *Tfad2c* depletion in any of the examined cell types. It is therefore unlikely that gRNAs are targeting sites within CREs which only act on *Tfad2c*, however it is still possible that significant gRNAs identified in this screen target sites within CREs which may also regulate *Tfad2c*.

Outside of the *Tfad2c* TAD and the surrounding regions, only three other tiled regions were



**Figure 39: Genomic features of significant gRNAs.** Pie charts of genomic features identified in significantly negatively- or positively-enriched gRNAs per cell type are indicated. A features key is included below, and “no features” indicates broadly intergenic sequences exhibiting no overlap with the listed features.

detected in the screen: potential regulatory region 1 (PRR1; chr2:170,246,202-170,503,872), positive control enhancer 3 (PCE3; chr14:99,323,583-99,327,543), and negative control enhancer 2 (NCE2; chr14:99,333,716- 99,341,806). 41 significant gRNAs were targeted to unique sites in PRR1 (data not shown), and thus interesting potential targets for further characterization may be found in this region, which was initially included in the screen because it contained sites identified via Hi-C to be directly interacting with the *Tfap2c* promoter. Only one significant gRNA was detected in PCE3, although it was detected in both negatively enriched ESCs and positively enriched PGCs, and only one significant gRNA was detected in NCE2. Aside from PCE3, no gRNAs targeting other positive control enhancers identified to be necessary for ESC survival by Kearns *et al.* [267] were shown to be enriched in any sample. It may thus be concluded that this screen could not detect entire enhancers. However, due to the identification of significant gRNAs in intronic or intergenic CREs, it is very possible that the screen was able to identify TF binding sites within enhancers which may be further characterized.

Taken together, this screen appeared was only partially successful. I was able to identify most PGC-gene targeting gRNAs, although it was not able to identify *Tfap2c*-targeting



**Figure 40: The *Aurka/Cstf1* locus on chromosome 2 indicates negative enrichment of significant gRNAs in exons following EpiLC induction.** Significant gRNAs are plotted in IGV and demonstrate enrichment of exon-targeting gRNAs in the EpiLC negative enrichment sample. WT RNA-seq and ATAC-seq data are included and labeled for reference, and chromosomal coordinates and increments are noted. Image generated in IGV.

gRNAs. It is unlikely that significant gRNAs identified in this screen are thus targeting CREs which only act on *Tfap2c*, however the possibility remains identified gRNAs may target CREs which also act on *Tfap2c*. Interestingly, gRNAs targeting individual exons of the genes *Aurka*, *Cstf1*, and *Rae1* were determined to be significantly negatively enriched in EpiLCs, indicating a potential role for these gene in EpiLC induction or differentiation. Finally, many significant gRNAs were detected in regions annotated to cCREs, indicating that the gRNAs may have targeted TF binding sites relevant to cell type-specific CREs. Ultimately, although the primary objective of the screen was not met, many interesting gRNAs may nevertheless be investigated to identify CREs required for ESC, EpiLC, or PGCLC fate.





# Discussion

AP2 $\gamma$  is a truly remarkable transcription factor required in three key moments of early embryonic development: blastocyst formation [126], TE induction [70], and PGC differentiation [35]. Accordingly, the KD of maternal and zygotic *Tfap2c* mRNA results in embryonic death by E3.5 due to failure of blastocyst formation [126], while genetic KO of AP2 $\gamma$  results in embryonic lethality between E8.5-E10.5 [124]. The study of AP2 $\gamma$  is therefore vital for understanding embryonic development as well as PGC differentiation. The work described in this dissertation aimed to improve our understanding of the role of AP2 $\gamma$  in early development, as mimicked by an *in vitro* system, in two main respects: 1) investigating the role of AP2 $\gamma$  as a regulator of gene expression and chromatin accessibility by characterizing AP2 $\gamma$ -KO cell lines via RNA-seq and ATAC-seq, and 2) employing a CRISPR/Cas9-based tiling screen of the TAD harboring AP2 $\gamma$ -encoding gene *Tfap2c* to regulatory elements acting on the gene. Taken together, these two projects were designed to broaden our knowledge concerning regulation by and of AP2 $\gamma$  in the process of PGCLC differentiation. Ultimately, the RNA-seq and ATAC-seq analyses demonstrated the misregulation of genes and differential peak accessibility throughout all examined cell states, and raises interesting questions about the role of AP2 $\gamma$  as a transcriptional regulator and potential pioneer factor. Further, the tiling screen identified sites of interest which may be located within regulatory elements required for the maintenance or induction of specific cell types. The findings, interpretations, and outlook of both projects are discussed in detail in the following sections.

## 4.1 The role of AP2 $\gamma$ in PGC induction

AP2 $\gamma$  has been well-established as a PGC factor in the mouse [37; 50; 35], however it is the least examined of the three key PGC factors AP2 $\gamma$ , BLIMP1, and PRDM14. Published work has established that conditional germ cell-specific TNAP-Cre AP2 $\gamma$ -KO mouse lines result in the birth of healthy but sterile mice [35]. In this mouse model, PGC specification occurs and PGCs are identifiable by E8.0 but are lost by E12.5, and the gonads of the resulting adult mouse are reduced in size [35]. *In vitro*, AP2 $\gamma$ -KO PGCLCs have been analyzed by microarray transcriptomic analysis, which demonstrated reduced expression of PGC markers *Dazl*, *Dppa3*, and *Kit*, as well as upregulation of the primed pluripotency marker *Dnmt3b*, indicating misregulation of the PGC gene program in the absence of *Tfap2c* expression [50]. While published literature shows that AP2 $\gamma$  is a factor required for PGC differentiation, open questions remain regarding how AP2 $\gamma$  exerts the transcriptional

control necessary to ensure PGC fate. One aim of my work was to improve our understanding of AP2 $\gamma$  as a transcriptional regulator by investigating WT and AP2 $\gamma$ -KO cells via RNA-seq and ATAC-seq, to identify potential misregulation and alterations in chromatin accessibility as ESCs embark on the developmental pathway towards PGCLC fate.

#### 4.1.1 Loss of AP2 $\gamma$ affects the morphology of ESCs, but not that of EpiLCs or EBs

To begin this work, AP2 $\gamma$ -KO cell lines were generated, validated, and characterized in both B1 and SGET ESC lines (Figures 8 and 20). The B1-based AP2 $\gamma$ -KO ESC lines grew normally in culture and were largely phenotypically normal, however single, differentiated cells were seen in greater numbers around the periphery of rounded ESC colonies (Figure 8). Interestingly, a recent study showed that single mouse ESC colonies grown on artificial mesh islands demonstrate alterations in gene expression based on the location of the cell relative to the overall colony, *i.e.* within a single colony, cells grown in the center strongly expressed OCT4 and expressed low levels of AP2 $\gamma$  and BLIMP1, while the cells on the periphery of the colonies strongly expressed AP2 $\gamma$  and BLIMP1 and OCT4 expression decreased [271]. It is further worth noting that tight-junction (TJ) proteins have been identified as key components in serum-free hESC colony formation [272], and that AP2 $\gamma$  has been identified as a regulator of TJ assembly in the blastocyst [126]. Thus, it can be speculated that AP2 $\gamma$  expression may be involved in the maintenance of colony borders, and in its absence peripheral cells of colonies may be more likely to undergo differentiation. To better understand if the results demonstrated by Ando *et al.* may be found in 2i+LIF culture conditions and potentially explain the observed phenotype, detailed confocal imaging of B1 WT stained colonies could be performed to investigate if an AP2 $\gamma$ /BLIMP1 localization-specific expression phenotype exists in 2i+LIF culture conditions. If so, B1 WT and KO GFP $^{+/-}$  ESCs could be FACS-sorted and analyzed via qPCR for pluripotency, cell junction, and motility factors to potentially identify changes in gene regulation resulting in the phenotype identified in AP2 $\gamma$ -KO cells.

*Tfap2c* has also been identified as a potential regulator of cellular locomotion: one study found that KD of *Tfap2c* in human fibroblasts resulted in failure of the fibroblasts to correctly reorient themselves following collision, leading to the generation of a misaligned extracellular matrix [273]. KD of *Tfap2c* in human breast cancer cells induced epithelial to mesenchymal transition (EMT) [76], a process often upregulated in cancer through which epithelial cells adopt a mesenchymal phenotype and increase their migratory behavior [274]. It is therefore possible that the single, differentiated cells identified at the periphery of AP2 $\gamma$ -KO ESC lines exhibit a migratory phenotype. This could be investigated by live cell

imaging to observe if and how the single cells travel *in vitro*, or via scratch-wound assay to quantify the speed at which cells repair a scratched colony.

Understanding the effect of AP2 $\gamma$ -KO on cell motility is of interest as a necessary event during PGC development is the migration from the proximal-posterior epiblast [26] to the genital ridge [41]. AP2 $\gamma$  expression has been identified in migrating PGCs, and KO of AP2 $\gamma$  has resulted in the failure of PGCs to migrate [35]. Thus, improved understanding of the role of AP2 $\gamma$  in cellular locomotion may improve our understanding of PGC migration. Further, AP2 $\gamma$  has been implicated as a factor in numerous cancers (see Section 1.3.1), and was also shown to upregulate cancer markers in this work (Figure 18). EMT is a feature of many cancers [274], and better understanding of how AP2 $\gamma$  may contribute to the EMT phenotype or to altered cellular motility more generally may shed light on its complicated and context-dependent role as either an ameliorating or exacerbating factor in cancer pathogenicity.

Beyond the phenotype described for AP2 $\gamma$ -KO ESCs, the morphology of AP2 $\gamma$ -KO and WT EpiLCs could not be distinguished, as was also observed between AP2 $\gamma$ -KO and WT EBs. It is of note that the KO of AP2 $\gamma$  did not appear to reduce the percentage of cells expressing BLIMP1 in d4/d6 +cyto EBs (Figures 9-10, 23). This is consistent with previous studies showing that BLIMP1 is activated upstream of AP2 $\gamma$  during PGC induction and acts directly on *Tfap2c* to induce its expression [26; 34; 37]. However, another study showed that KO of AP2 $\gamma$  in ESCs resulted in reduced BLIMP1 expression upon PGCLC induction [50].

To explore the effect of the KO of AP2 $\gamma$  beyond the morphological phenotype, RNA-seq and ATAC-seq were performed on WT and AP2 $\gamma$ -KO ESCs, EpiLCs, and d4/d6 GFP+ cells to identify potential changes in the transcriptome or genomic accessibility induced by the KO of AP2 $\gamma$ , and ultimately to interpret what differences may mean for the individual cell state as well as the functional effect it may have during or following PGCLC induction. The effect of AP2 $\gamma$ -KO on ESCs will be discussed first, as these cells represent the first step in the developmental pathway leading ultimately to PGC fate.

#### **4.1.2 AP2 $\gamma$ -KO results in misregulation of imprinted genes and dramatically alters chromatin accessibility**

Although AP2 $\gamma$  is not considered a pluripotency factor in the mouse [257], 449 genes were significantly differentially expressed (DE) in the AP2 $\gamma$ -KO ESCs relative to the WT. This included the downregulation of naïve pluripotency factors *Dppa3* and *Kit* and the upregula-

tion of *Jam2* and *Klf2* (Figure 13C). *Dppa3* represses the maintenance of DNA methylation resulting in global DNA hypomethylation of the ESC genome [275]. In addition, it counterintuitively protects methylation of largely maternally imprinted genes [276]. Imprinted genes are a class of genes wherein either the maternal or paternal allele is silenced and expression is repressed, which leads to mono-allelic gene expression [277]. The downregulation of *Dppa3* may thus explain the misregulation of twelve imprinted genes in the KO ESCs (Figure 13D). *Dlk1* (Delta like non-canonical Notch ligand 1) is a paternally imprinted gene which has been identified to be significantly upregulated in AP2 $\gamma$ -KO ESCs, and is known to be directly protected from demethylation by *Dppa3* [278]. Misregulation of imprinting has been shown to affect ESC self-renewal and differentiation capacity [277], and it has been proposed that imprinted genes should be observed as markers of ESC culture quality, with misregulation indicating poor developmental potential [279]. Thus, in the ESCs both imprinted genes and an important regulator of imprinting are misregulated. This raises the question of whether AP2 $\gamma$ -KO ESCs exhibit the same developmental potential as WT ESCs.

Additionally, a small number of genes associated with specific extra-embryonic and proper embryonic lineages, including the PE, TE, and mesoderm, were upregulated in the AP2 $\gamma$ -KO ESCs relative to WT (Figure 13C). However, as AP2 $\gamma$ -KO ESCs may be maintained in cell culture and demonstrate normal self-renewal, it is unlikely that the differential expression observed in AP2 $\gamma$ -KO ESCs indicates that the cells have adopted another extra- or proper embryonic fate. Rather, in concordance with the G0seq gene ontology analysis of DE AP2 $\gamma$ -KO ESC genes indicating numerous developmentally-associated terms (Figure 13B), it is possible that the AP2 $\gamma$ -KO ESCs become more heterogeneous in culture and thus more likely to spontaneously differentiate, potentially resulting in the phenotype discussed in Section 4.1.1.

Following RNA-seq analysis, ATAC-seq was performed to identify whether the KO of AP2 $\gamma$  resulted in differentially accessible chromatin, but also to determine if a correlation could be found between identified DE genes and differentially accessible peaks (DAPs). ATAC-seq analysis demonstrated a dramatic effect upon chromatin accessibility in the AP2 $\gamma$ -KO ESCs: 8.3% of all investigated peaks were considered DAPs, with 4.2% (6,500/156,620 peaks) positively enriched and 4.1% (6,463/156,620 peaks) negatively enriched (Figure 28A). AP2 $\gamma$  was the third most negatively enriched motif in the KO ESCs, while AP2 $\alpha$  was the second-most negatively enriched motif (Figure 29A). This result demonstrates that in the absence of AP2 $\gamma$  in ESCs the AP2 $\gamma$  motif is significantly less accessible, and strongly indicates that AP2 $\gamma$  might act as a pioneer factor in ESCs. Alternatively, it is possible that AP2 $\gamma$  regulates other pioneer factors which make these regions accessible, and in AP2 $\gamma$ -KO

lines the misregulation of these factors leads to reduced AP2 $\gamma$ -motif accessibility.

Numerous SOX TF family motifs were found in the ten most negatively enriched motifs of AP2 $\gamma$ -KO ESCs, including the motif of naïve pluripotency factor SOX2. SOX family TFs are pioneer factors [280], and it is intriguing that in the absence of AP2 $\gamma$ , SOX family motifs are negatively enriched, especially given that no SOX factors were found to be DE in AP2 $\gamma$ -KO ESCs. The potential implications of this result are discussed further in Section 4.1.5. TEAD family factor motifs were found to be positively enriched in the AP2 $\gamma$ -KO ESC and d4 GFP+ cell DAPs (Figure 29B, 31B). TEAD family TF factors are involved in neural crest and trophectoderm development and are also implicated in tumorigenesis [281]. It is intriguing that in AP2 $\gamma$ -KO ESC RNA-seq data, no TEAD factor genes were found to be DE, however in the AP2 $\gamma$ -KO d6 GFP+ cell RNA-seq data, TEAD factor genes *Tead1*, *Tead2*, and *Tead3* are upregulated (Appendix 7.9.3.3). It was shown that KD of AP2 $\gamma$  in mouse ESCs resulted in the differentiation of neural cell types in embryoid bodies (EBs) [262]. Additionally, human *TFAP2C*<sup>-/-</sup> cells have been shown to upregulate markers of neural lineage [257]. Thus, the enrichment of neural crest-associated TEAD factors may prime the ESCs towards later expression of neural factors (Figure 58), and as such is consistent with previously published works and helps to validate the output of the ATAC-seq analysis.

The annotation of both negatively and positively enriched DAPs in AP2 $\gamma$ -KO ESCs resulted in the identification of over 8,000 genes, however very few corresponded to genes found to be DE in the AP2 $\gamma$ -KO ESC RNA-seq data; only 2.6% (107/4,076) of genes in positively enriched DAPs and only 2.9% (128/4,395) of genes found in negatively enriched DAPs were found to be DE (Appendix 7.9.1.7). Most ATAC-seq peaks are localized to enhancers [255], and thus an identified DAP may not have a functional association with the nearest TSS of a gene; for example, an *Shh* (sonic hedgehog) enhancer was identified over 1 megabase (Mb) away from the target gene [139], and automated annotation of such a site would almost certainly annotate the enhancer to a closer, unrelated gene [221]. Thus, I examined the relationship between DAPs annotated to promoters and DE genes identified in the RNA-seq dataset, however I again found very few DAPs corresponded to DE genes (Appendix 7.9.1.7), with the exception of *Dppa3*. As *Dppa3* is downregulated in the ESCs, it is thus likely that AP2 $\gamma$  directly binds to the promoter of *Dppa3* in ESCs and regulates its expression. This may explain the misregulation of imprinted genes observed in AP2 $\gamma$ -KO ESCs, and should be confirmed by identifying the genomic binding sites of AP2 $\gamma$  in ESCs with CHIP-seq (Chromatin immunoprecipitation sequencing) or CUT&RUN (Cleavage under targets and release using nuclease). If AP2 $\gamma$  binds directly to the *Dppa3* promoter, it would place AP2 $\gamma$  within a network of factors required to maintain naïve pluripotency, and indicate AP2 $\gamma$  as

an upstream regulator of imprinting. More broadly, the ATAC-seq analysis of AP2 $\gamma$ -KO ESCs demonstrates dramatic changes in chromatic accessibility while the overall effect on gene expression is minor. This raises the possibility that the KO of AP2 $\gamma$  in ESCs may primarily affect the accessibility of enhancers and thus the developmental potential of the cells, without directly affecting gene expression within the ESCs.

### 4.1.3 AP2 $\gamma$ -KO EpiLCs upregulate lineage non-specific markers

*Tfap2c* is downregulated in EpiLCs (Figure 12) and is not expressed in post-implantation epiblast cells *in vivo* [123]. It was therefore surprising to find that more genes were considered significantly DE in EpiLCs (792; Figure 14) than in ESCs (449; Figure 13). This result may be explained by the fact that the AP2 $\gamma$ -KO ESCs demonstrated the misregulation of numerous genes, as well as differential accessibility of thousands of individual genomic loci. Thus, the increased number of misregulated genes identified in EpiLCs may derive from the misregulation observed in ESCs.

I found that genes associated with pluripotency and embryonic lineages were misregulated in EpiLCs, including the naïve pluripotency markers *Prdm14* and *Dppa3* (Figure 14C). Additionally, markers of formative pluripotency fate *Etv5* and *Fgf5* were also downregulated in the AP2 $\gamma$ -KO EpiLCs (14C). A small selection of markers indicating ectodermal, PE, and VE fate were also upregulated in the AP2 $\gamma$ -KO EpiLCs relative to WT (14D), and eleven imprinted genes were misregulated (14E). This raises the possibility that the AP2 $\gamma$ -KO EpiLCs may not have completely achieved formative pluripotency.

ATAC-seq analysis shows that far fewer DAPs are identified in the AP2 $\gamma$ -KO EpiLCs relative to the number identified in the AP2 $\gamma$ -KO ESCs (Figure 28B). Motif enrichment analysis shows that KLF4 and EKLF motifs were negatively enriched in DAPs of AP2 $\gamma$ -KO EpiLCs (Figure 30A). EKLF is a TF which forms active chromatin hubs (ACHs) during erythropoiesis [259], while KLF4 is a naïve pluripotency factor [282]. Among the positively enriched motifs, I identified four motifs associated with the POU family of TFs (Figure 30B). This family of TFs includes OCT4, which is considered a pioneer factor [283] as well as a core pluripotency factor in ESCs [284]. Its expression is maintained in EpiLCs, albeit through the regulatory action of different enhancers and promoters [285]. Although the positive enrichment of OCT4 motifs may suggest that AP2 $\gamma$  plays a role in controlling the accessibility of OCT4 binding sites in EpiLCs, expression of *Pou5f1*, encoding OCT4, is not altered in AP2 $\gamma$ -KO EpiLCs relative to WT. Further, while changes in the accessibility of CREs acting on OCT4 have been identified in this work in AP2 $\gamma$ -KO ESCs (Appendix 7.9.1.6), none were identified in EpiLCs. It is thus difficult to come to a conclusion about

the enrichment of motifs in AP2 $\gamma$ -KO EpiLCs, however it could be that with a relatively small number of peaks analyzed, the identified enrichment lacks developmental relevance.

Ultimately, it is not possible to determine whether the KO of AP2 $\gamma$  is directly or indirectly affecting the differential expression of genes or changes in chromatin accessibility, as it is possible that the identified misregulation derives from the already-misregulated AP2 $\gamma$ -KO ESCs. In order to investigate whether the KO of AP2 $\gamma$  affects the phenotype of EpiLCs, a degron system of AP2 $\gamma$  may be employed to deplete AP2 $\gamma$  at specific time points during differentiation. This method functions by tagging an endogenous protein of interest with a protein tag that in turn targets the protein for proteolysis. Such systems are available for use, including the auxin-inducible degron (AiD) system or the dTag system [286]. Once endogenous AP2 $\gamma$  has been depleted, for example over the 42 hour course of EpiLC induction, EpiLCs could be harvested and analyzed for expression of naïve and formative pluripotency markers.

#### 4.1.4 Loss of AP2 $\gamma$ results in impaired PGCLC differentiation

The RNA- and ATAC-seq analysis of d4/d6 GFP+ AP2 $\gamma$ -KO cells demonstrates that PGCLC differentiation is impaired in the absence of AP2 $\gamma$ . RNA-seq analysis identified 1,009 genes to be DE between AP2 $\gamma$ -KO d4 GFP+ and WT, of which 622 were upregulated and 387 were downregulated. By d6, 4,813 genes were identified to be DE in the AP2 $\gamma$ -KO GFP+ cells, of which 2,644 were upregulated and 2,169 were downregulated. Thus, the number of total DE genes from d4 to d6 of differentiation increased four-fold within two days. This increased misregulation is consistent with published work which shows that the conditional KO of AP2 $\gamma$  does not impair the specification of PGCs by E8.0 but rather impairs their differentiation and maturation, resulting in depletion of PGC populations by E12.5 [35]. In this work, the loss of AP2 $\gamma$  resulted in the downregulation of pluripotency and PGC genes (Figures 15C-17A,B), while numerous markers of somatic and extra-embryonic lineages, including the DE, ExE, mesoderm, mesendoderm, PE, TE, and VE were upregulated in both d4 and d6 datasets (Figures 15D-17C). As previously observed in the AP2 $\gamma$ -KO ESCs and EpiLCs, imprinted genes are also misregulated in the AP2 $\gamma$ -KO d4/d6 GFP+ cells (Figures 15E-17D). It may be concluded that the loss of AP2 $\gamma$  impairs the differentiation of PGCLCs and rather differentiates the cells towards somatic lineages.

I was able to confirm some RNA-seq gene expression results identified in the B1-based AP2 $\gamma$ -KO d6 GFP+ cells in the S-KO d6 GFP+/tdTomato- cells (Figures 21-23). However, *Tfap2a* was not upregulated but rather downregulated in the S-KO d6 GFP+/tdTomato- cells. While this result is intriguing, it must be acknowledged that the SGET-based AP2 $\gamma$ -

KO cell lines represent a different population of cells. The B1-based AP2 $\gamma$ -KO GFP+ d4/d6 cells represent all cells which have upregulated the PGC master regulator *Prdm1* and initiated the PGCLC transcriptional program, while upregulation of PGC downstream marker *Stella* indicates cells of a more established PGCLC fate. It may be of interest in the future to perform RNA-seq on S-KO d4/d6 GFP+/tdTomato- cells to identify if transcriptomic changes, for example the misregulation of imprinted genes, are maintained.

Gene ontology analysis of d4/d6 GFP+ cells shows that DE genes in AP2 $\gamma$ -KO cells are largely associated with developmental terms (Figures 31B-16B). These correspond to a substantial number of early embryonic lineage genes. The misregulation of non-PGC somatic markers raised the question of whether the d4/d6 GFP+ AP2 $\gamma$ -KO cells had entered a general state of global transcriptomic misregulation, or whether they had entered a lineage-specific transcriptional program. Interestingly, numerous markers of the embryonic multipotent progenitor (EMP) fate were upregulated in the d6 GFP+ AP2 $\gamma$ -KO cells, as well as markers which identify both EMP and basal cell (BC) fate. However, relatively few markers shared between EMP and luminal cell (LC) fate were upregulated (Figure 19). This is noteworthy, as the conditional mammary-specific knockout of *Tfap2c* in mice also resulted in an increased BC population and reduced LC population [76]. Thus, one possibility is that the generation of EBs in BMP4+ medium and in the absence of AP2 $\gamma$  drives the induction of an EMP-like signature in cells throughout the EB. In support of this theory, BMP4 is released by the ventral epidermis surrounding the mammary mesenchyme by at least E13.5 [287]. The mammary mesenchyme responds by upregulating BMP receptors and BMP4 targets of regulation, including the genes *Msx1* (Msh homeobox 1) and *Msx2* (Msh homeobox 2) [287]. All BMP receptor genes are upregulated in the d6 GFP+ AP2 $\gamma$ -KO cells relative to WT, as are *Msx1*, *Msx2*, and all *Id* (inhibitor of DNA binding) genes, which are also direct targets of BMP4 signaling [288]. In principle, this idea could be tested by generating EBs directly from AP2 $\gamma$ -KO ESCs in medium containing BMP4, and performing qRT-PCR on resulting EBs to identify expression of EMP markers such as *Sox11*.

Another possibility is that the EMP-like cells require both BLIMP1 expression and BMP4 signaling as well as the absence of AP2 $\gamma$  in order to correctly specify. However, the previously described BLIMP1+ mammary subpopulation is identified by E17.5 in the mouse [289], much later than the E9.5 identity of d6 GFP+ PGCLCs [24] or the E12.5 identity of EMPs [252] and does not express basal markers [289]. The contribution of BLIMP1 to the EMP-like cells of the d6 GFP+ AP2 $\gamma$ -KO cells may be easily tested, namely by knocking-out *Prdm1* (encoding BLIMP1), generating EpiLCs and subsequently EBs with BMP4+ medium, and analyzing the EBs on d6 to determine of markers of EMP fate. Additionally,



*Prdm1* may be overexpressed in WT and AP2 $\gamma$ -KO ESCs to identify if markers of EMP fate are upregulated. Alternatively, the upregulation of *Tfap2a* identified in the AP2 $\gamma$ -KO d4/d6 GFP<sup>+</sup> cells drives the EMP-like gene program in the presence of BMP4 and the absence of AP2 $\gamma$ . As with BLIMP1, this explanation falls short as, although AP2 $\alpha$  has been identified in mammary tissue [290], it has also been identified as a positive regulator of LC fate [290] and thus does not explain the upregulation of BC markers identified in the EMP-like cells generated in this work. The role of AP2 $\alpha$  in the generation of the EMP-like AP2 $\gamma$ -KO d6 GFP<sup>+</sup> cells may be tested as described with BLIMP1; the gene could be knocked out and the cells differentiated towards PGCLC fate as previously described and markers could be analyzed via qRT-PCR, or AP2 $\alpha$  may be overexpressed in ESCs to identify if EMP markers are correspondingly upregulated.

Ultimately, the question raised by the aberrant gene expression observed in AP2 $\gamma$ -KO EpiLCs remains: is the failure of PGCLCs to specify a consequence of the absence of AP2 $\gamma$ , or does it arise from misregulation in the ESCs or EpiLCs? Published work demonstrating that PGC-specific AP2 $\gamma$ -KO lines demonstrate the loss of PGCs, resulting in healthy albeit sterile mice [35], indicates that indeed AP2 $\gamma$  is required for PGC induction. Nevertheless, the specific effects upon the transcriptome of AP2 $\gamma$ -KO d6 GFP<sup>+</sup> cells may be different than those observed in PGC-specific KO mice. To determine if dramatic transcriptomic changes observed in the AP2 $\gamma$ -KO d4/d6 GFP<sup>+</sup> cells, including misregulation of imprinting genes and upregulation of EMP genes, are fundamental to the KO of AP2 $\gamma$  or are the ‘by-product’ of AP2 $\gamma$ -KO ESC/EpiLC misregulation, the targeted degradation of AP2 $\gamma$  may be employed. Depletion of AP2 $\gamma$  at the moment of PGC induction and subsequent qRT-PCR of meGFP<sup>+</sup> cells collected from d4/d6 EBs for identified DE markers of imprint misregulation including *Peg13* and *H19*, or previously mentioned EMP markers, should allow for this point to be clarified.

#### 4.1.5 Chromatin accessibility is altered during PGCLC induction from AP2 $\gamma$ -KO EpiLCs

As RNA-seq analysis demonstrated global transcriptomic misregulation in the AP2 $\gamma$ -KO d4/d6 GFP<sup>+</sup> cells, it was of interest to determine whether the KO of putative pioneer factor AP2 $\gamma$  resulted in a chromatin accessibility phenotype. As the quality of the WT d6 GFP<sup>+</sup> samples varied (Figures 7.9.1.4, 7.9.1.2D), few DAPs were identified and these results will not be discussed. In the AP2 $\gamma$ -KO d4 GFP<sup>+</sup> cells, 0.18% (161/91,003) of the identified DAPs were negatively enriched while only 0.077% (70/91,003) of the DAPs were positively enriched. This represents a very small portion of all peaks and stands in contrast to the much larger change in chromatin accessibility identified in AP2 $\gamma$ -KO ESCs. Thus,

it may be that AP2 $\gamma$  plays a context-dependent role in ESCs and PGCLCs.

AP2 $\alpha$  and AP2 $\gamma$  were the two most negatively enriched identified motifs in the AP2 $\gamma$ -KO d4 GFP $^+$  cells, and numerous KLF family motifs were also found to be negatively enriched (Figure 31). Within the positively enriched motifs identified in the AP2 $\gamma$ -KO d4 GFP $^+$  cells were numerous SOX family factors, as well as the primed pluripotency factor OTX2, two TEAD family factors, and two ZIC family factors. It is striking that SOX family factors were identified to be negatively enriched with the AP2 factors in the AP2 $\gamma$ -KO ESCs (Figure 29), and in the d4 GFP $^+$  AP2 $\gamma$ -KO cells the opposite effect may be observed. The upregulation of SOX and ZIC family factors was identified by Pastor *et al.* in *TFAP2c* $^{-/-}$  enhancers following failed reversion from primed to naïve human ESC states, and they determined that the positive enrichment of these factors is an indication of neural fate [257]. This corresponds to the upregulation of neural stem cell DE genes identified in d4/d6 GFP $^+$  AP2 $\gamma$ -KOs (Figure 58). A likely explanation for the negative enrichment of SOX factors in AP2 $\gamma$ -KO ESC DAPs and subsequent positive enrichment in AP2 $\gamma$ -KO d4 GFP $^+$  cells is simply upregulation of *Sox* factor expression in the AP2 $\gamma$ -KO d4/d6 GFP $^+$  cells relative to the AP2 $\gamma$ -KO d4 ESCs (Appendix 7.9.3.2). As SOX family factors are pioneer factors [280], their upregulation intuitively corresponds to the increased chromatin accessibility of their motifs.

Annotation of DAPs resulted in 147 genes annotated to the negatively enriched DAPs of which 22 corresponded to DE genes (15.0%), and 67 genes annotated to the positively enriched DAPs of which 8 corresponded to DE genes (11.9%; Appendix 7.9.1.7). No DAPs in the positively enriched dataset were annotated to promoters, while four of the DAPs in the negatively enriched data were annotated to promoters, of which three corresponded to the DE downregulated genes *Tfap2c*, *Mest*, and *Peg13*. *Mest* and *Peg13* are imprinted genes whose DAPs were depleted in all AP2 $\gamma$ -KO cells analyzed via ATAC-seq (Appendix 7.9.1.6). *Mest*-KO mice exhibit a reduction of mesodiencephalic dopaminergic neurons and a behavioral phenotype reminiscent of Parkinson’s disease [291], while *Peg13*-KO mice exhibit behavioral changes including depression [292]. As AP2 $\gamma$  has been shown to be expressed in the brain [293; 78], it is intriguing that the KO of AP2 $\gamma$  results in DE expression as well as negatively enriched DAPs of genes associated with behavioral phenotypes.

#### 4.1.6 Different effects of AP2 $\gamma$ -KO on ESCs and d4 GFP $^+$ cells

One unresolved issue identified this work is the counterintuitive result presented by the RNA-seq and ATAC-seq datasets when comparing AP2 $\gamma$ -KO ESCs and d4 GFP $^+$  cells.

Namely, that AP2 $\gamma$ -KO ESCs exhibit less DE gene expression but numerous DAPs while d4/d6 GFP+ AP2 $\gamma$ -KO cells demonstrate dramatic DE gene expression but exhibit far fewer DAPs. A speculative explanation may be that a redundant pioneer such as AP2 $\alpha$  partially compensates for the loss of AP2 $\gamma$  on d4/d6 of PGCLC differentiation. The question of whether AP2 $\alpha$  and AP2 $\gamma$  are redundant has been studied in the early murine embryo [88], in the zebrafish neural crest [107], and in murine epidermis [106]. The resulting double KO of both factors often demonstrates a more severe phenotype, *e.g.* KO of both AP2 $\alpha$  and AP2 $\gamma$  in the mouse results in earlier embryonic lethality than either factor alone [88]. However, neither factor is truly redundant for the other, meaning the continued expression of AP2 $\alpha$  is not sufficient to rescue AP2 $\gamma$ -KO cells and vice versa. This is consistent with my data, as *Tfap2a* expression is upregulated in d4/d6 GFP+ AP2 $\gamma$ -KO cells however does not rescue the PGC gene program (Figure 18). A possible explanation for this is that AP2 $\gamma$  and AP2 $\alpha$  could be functionally redundant as pioneer factors but not as transcriptional regulators.

In this model, the upregulation of AP2 $\gamma$  in AP2 $\gamma$ -KO d4/d6 GFP+ cells leads to the maintained accessibility of chromatin bearing AP2 $\gamma$ /AP2 $\alpha$  motifs and thus explains why so few DAPs may be identified in the d4 GFP+ AP2 $\gamma$ -KO cells. However, as AP2 $\alpha$  cannot compensate for the transcriptional regulation of AP2 $\gamma$ , widespread transcriptional misregulation occurs. In support of this theory is the fact that both AP2 factors share a conserved AP-2 binding domain [97], which detects extremely similar motifs (see for example Figure 27A). Additionally, they are both putative pioneer factors [52; 73; 77; 102; 103]. Further, while the AP-2 binding domain is conserved between the two proteins, only 52% of the overall protein sequence is conserved [101]. This might explain why the two proteins exhibit different interaction partners and tissue-specific expression [79].

In opposition to this theory is the fact that AP2 $\alpha$  binding motifs are even more negatively enriched in the d4 GFP+ AP2 $\gamma$ -KO cells than AP2 $\gamma$  binding motifs (Figure 31A). However, it should be noted that this enrichment analysis was only performed upon a small number of DAPs, and thus it is possible that the negative enrichment of AP2 $\alpha$ /AP2 $\gamma$  corresponds specifically to peaks requiring AP2 $\alpha$ /AP2 $\gamma$  heterodimers, which have been identified in the chick neural crest [77]. In total, 79 negatively enriched DAPs were found to contain the AP2 $\alpha$  binding motif, while 87 contained the AP2 $\gamma$  binding motif, meaning almost all DAPs contained the motifs of both factors. To examine the potential functionally pioneer factor redundancy of AP2 $\alpha$  and AP2 $\gamma$ , ChIP-seq or CUT&RUN libraries of both factors could be generated to identify unique and overlapping binding sites. AP2 $\alpha$  binding sites in AP2 $\gamma$ -KO cells could then be identified, to determine whether the upregulation identified in AP2 $\gamma$ -KO d4/d6 GFP+ cells results in increased AP2 $\alpha$  binding in AP2 $\gamma$  motif-containing

peaks. Further, as *Tfap2a* was not found to be upregulated in the S-KO d6 GPF+ cells (Figure 23), these cells could be used for ATAC-seq to determine if more DAPs may be identified when *Tfap2a* is not upregulated.

#### 4.1.7 Conclusions on the study of AP2 $\gamma$ activity

In this work, I have clearly demonstrated that PGCLC development from AP2 $\gamma$ -KO ESCs is impaired. However, I have also identified new and intriguing potential targets of AP2 $\gamma$  regulation, namely imprinted genes and EMP genes. I have raised questions about the true identity of ESCs and EpiLCs following AP2 $\gamma$ -KO which should be investigated to better determine the role the protein plays on early development. I have also proposed a model of AP2 $\alpha$  as functionally redundant to AP2 $\gamma$  within the context of chromatin accessibility and non-redundant in the context of gene regulation. To conclude this first aim of my work, concerning the activity of AP2 $\gamma$  as a transcriptional regulator and potential pioneer factor, I will discuss the conclusions I have come to on AP2 $\gamma$ , as can best be determined by the work I have presented here.

##### 4.1.7.1 AP2 $\gamma$ should be evaluated as a potential factor supporting naïve pluripotency

In this work, I have demonstrated that the loss of AP2 $\gamma$  results in altered gene expression and chromatin accessibility. Recently, it was shown that AP2 $\gamma$  motifs are enriched in human ESC (hESC) naïve-specific ATAC peaks [257]. *TFAP2C*-KO primed hESC lines, which when maintained in the primed state, did not show changes in pluripotent marker expression. However, naïve pluripotency markers were downregulated and naïve-specific DAPs were no longer accessible upon reversion of *TFAP2C*-KO hESCs to the naïve pluripotent state, which was less pronounced when analyzing mouse ESCs. This study concluded that “AP2 transcription factors play a more modest role in murine than human naïve states” [257]. The effect I observed in the mESCs was much larger than what is described by Pastor *et al.*, as I identify almost 13,000 DAPs in the AP2 $\gamma$ -KO ESCs relative to WT. The differences observed between these works may be attributed to several factors: 1) they compared different mESC lines to so-called mouse epiblast stem cells (EpiSCs), 2) they used different bioinformatics tools, and 3) they did not examine the ATAC-seq profiles of *Tfap2c*-KO cells alone but only AP2 $\gamma$ <sup>-/-</sup>/AP2 $\alpha$ <sup>-/-</sup> dual KOs.

It is possible that AP2 $\gamma$  is required for the correct maintenance or priming of the mESC naïve pluripotent state, while nevertheless playing a smaller and less consequential role than is observed in hESCs. In this proposed model, AP2 $\gamma$  may directly and indirectly regulate genes affecting chromatin accessibility in the naïve state, including DNA methylation reg-

ulators/modifiers and pioneer factors which affect chromatin accessibility, however the KO of the protein is insufficient to result in reduced proliferation or complete differentiation of the ESCs. In support of this theory are the following facts: AP2 $\gamma$  is detectably expressed on an mRNA (Figure 12) and protein level (Figure 8) in mESCs, which correlates with published work demonstrating that it is expressed *in vivo* in the early mouse embryo [88]. AP2 $\gamma$ -KO ESCs demonstrate a phenotype as well as differential expression of 449 genes and differential accessibility of almost 13,000 chromatin peaks, indicating significant transcriptional and chromatin accessibility consequences for AP2 $\gamma$ -KO ESCs. Further, AP2 $\gamma$ -KO ESCs downregulate *Dppa3*, and ATAC-seq peaks associated with the *Dppa3* promoter are negatively enriched. Additionally, *Dnmt3l* is found to be upregulated, and the promoter of *Dnmt3l* is identified as a positively enriched DAP in the AP2 $\gamma$ -KO ESCs. *Dnmt3l* is a *de novo* DNA methylation enzyme, and the misregulation of the genes *Dppa3* and *Dnmt3l* is associated with negative enrichment in genome accessibility due to DNA methylation. Additionally, several putative pioneer factors including *Pbx1* (PBX homeobox 1), *Ets1* (ETS proto-oncogene 1), *Meis2*, and *Meis3* (meis homeobox 3) as well as histone linker protein *H1f0* (Histone 1.0 linker)<sup>3</sup> were found to be upregulated in the AP2 $\gamma$ -KO ESCs, while chromatin-remodeling or accessibility associated proteins *Tet1* (Tet methylcytosine dioxygenase 1), *Hira* (Histone cell cycle regulator), and *Chd3* (Chromodomain-helicase-DNA-binding protein 3) were downregulated. Of particular note are *H1f0*, which associates with transposable elements and repetitive, non-coding, intergenic satellite regions and whose upregulation is associated with differentiation [295], and *Hira*, which assists in transition of enhancers to the poised state [296]. It should be noted that, with the exception of *Pbx1*, *Meis2*, and *Meis3*, none of these genes are DE in the AP2 $\gamma$ -KO d4 GFP+ cells. The widespread misregulation of factors associated with chromatin remodeling, accessibility, and methylation may explain the similar number of DAPs found to be both positively and negatively enriched in the AP2 $\gamma$ -KO ESCs.

The low expression of *Tfap2c* in the ICM of the murine pre-implantation embryo is one of the grounds on which Pastor *et al.* discount the likelihood of AP2 $\gamma$  acting as a murine pluripotency factor. To counter this argument, an interesting protein which may be used as a case study for low but important expression in naïve ESCs is DAZL. DAZL is an RNA-binding protein which was generally perceived to be a germ-cell specific protein [297]. However, work by Welling *et al.* demonstrates that DAZL is expressed in 80% of mESCs in 2i+LIF culture and enhances the translation of *Tet1* mRNA, thus positively regulating global hypomethylation in ESCs [298].

<sup>3</sup>Although most histone genes do not have poly(A) tails and thus would not be isolated in this RNA-seq dataset, histone gene *H1f0* does in fact have a poly(A) tail [294] and thus was successfully sequenced.

In opposition to the idea that AP2 $\gamma$  may play a role in maintaining naïve pluripotency is the fact that Auman *et al.* found that mice with chimeric embryonic AP2 $\gamma$ -KO genotypes specific to the embryo proper developed normally [124]. It should be noted that the chimeric embryonic AP2 $\gamma$ -KO mice described by Auman *et al.* were not characterized beyond E12.5; the survival rate, fertility, behavior, and physical characteristics of adult chimeric AP2 $\gamma$ -KO mice are not indicated. This omission is notable, as tissue-specific AP2 $\gamma$ -KO mice have demonstrated impaired vision [293], cognitive impairment [78], reduced sex organ size [35], and infertility [35]. Depending on the gene of interest, chimeric mouse models may allow for compensation of an otherwise significant phenotype [299]. This fact is acknowledged in the Auman *et al.* paper, and as such they additionally developed a tetraploid chimeric embryo, in which the extraembryonic tissues are WT and the embryo proper is almost entirely comprised of AP2 $\gamma$ -KO cells [124]. They do not describe the resulting chimeric AP2 $\gamma$ -KO embryos beyond E12.5, and do not provide information about the phenotypes of the embryos in late gestation or after birth. As far as I can determine, this is the only paper describing exclusively proper embryonic AP2 $\gamma$ -KO mice, and as such it may be concluded that information is absent on the effect of AP2 $\gamma$ -KO on developmental potential and postnatal survival of mice.

In order to better understand the role AP2 $\gamma$  may play in naïve pluripotency, teratomas could be generated in mice with the KO1 or KO2 lines described in this paper. Teratomas, in the context of pluripotency research, are artificial tumors generated by injecting stem cells into the non-uterine tissue of immunocompromised mice [300]. If the resulting tumors give rise to the three germ layers of ectoderm, mesoderm, and endoderm, the injected cells were determined to be truly pluripotent [300]. However, the ethics of this experiment are under debate, and the method is controversial due to pain and suffering of the injected mice [300]. To overcome the ethical burdens of teratoma generation, a gastruloid could be generated from either KO1 or KO2 lines. Gastruloids are EBs generated from ESCs which are intended to mimic the process of gastrulation in the early embryo, and induce differentiation of ESCs towards neural, endodermal, and mesodermal fates [301]. WT and AP2 $\gamma$ -KO ESCs could thus be used to generate gastruloids, and the differentiation efficiency towards neural, endodermal, and mesodermal fates could be compared. Additionally, an *in vitro* teratoma assay has been developed for hESCs [302], and one could attempt to adapt this protocol to mESCs and compare the developmental potential of WT and AP2 $\gamma$ -KO *in vitro* teratoma EBs. Finally, the reversion experiment described by Pastor *et al.* in which primed hESCs are reverted to naïve fate [257] could also be applied to mouse cells, wherein AP2 $\gamma$ -KO EpiSCs could be reverted towards ESC fate [303] and ATAC-seq could be performed to identify if naïve peaks re-emerge.

#### 4.1.7.2 The characterization of AP2 $\gamma$ as a potential pioneer factor is not complete

With the work I have presented here, it is not possible to conclude whether AP2 $\gamma$  is a pioneer factor in the cell types I have studied. To continue the work of characterizing AP2 $\gamma$  as a potential pioneer factor, the next step is to perform ChIP-seq or CUT&RUN to identify where AP2 $\gamma$  directly interacts with chromatin throughout development. Additionally, ChIP-seq or CUT&RUN libraries of the histone modification H3K4me1 could be generated; H3K4me1 is not only a mark identified to flank primed or active enhancers [137], but also is found to colocalize with putative pioneer factors on closed chromatin [253]. Thus, sites where AP2 $\gamma$  may behave as a pioneer factor may be identified by looking for direct binding of AP2 $\gamma$  to closed chromatin colocalizing with H3K4me1.

By characterizing AP2 $\gamma$  as a pioneer factor and identifying its targets, one can better understand the process of differentiation between cell types. Further, misregulation of AP2 $\gamma$  has been implicated in cancer including breast cancer (see Section 1.3.1). Thus, our improved understanding of how and whether AP2 $\gamma$  opens chromatin will potentially improve our understanding of the protein in a pathogenic context.

## 4.2 Single gRNA tiling screen of *Tfap2c* TAD identifies single gRNAs of interest

To accomplish the second aim of this work, a CRISPR/Cas9-based tiling screen of the *Tfap2c* TAD was performed with the aim of identifying regulatory elements acting on the gene *Tfap2c*. 328 out of 29,192 (1.12%) unique gRNAs were considered significant. The output and outlook of the screen, as well as suggested improvements for future screens, are discussed in the following sections.

### 4.2.1 Basis of screen

The basis of the lentiviral pooled screen performed in this work was to tile a defined set of coordinates in the genome with single gRNAs, in this case the TAD of the gene *Tfap2c*, to identify *cis*-regulatory elements (CREs) acting *Tfap2c*. Lentiviral transduction allows for the random insertion of virally-packaged gRNA sequences into the genome, and the determination of the multiplicity of infection (MOI) allows for statistical confidence that most cells will only be infected with one viral particle [304]. As such, the Cas9-expressing SGET-Cas9 ESCs employed in this screen statistically were infected by one lentiviral particle and subsequently expressed one gRNA, thus generating a potential KO phenotype specific to the gRNA they expressed. Should the gRNA target Cas9 to a TF binding site affecting cell

survival or proliferation, or in the case of cells sorted via FACS the loss of fluorescence, the depletion of that cell from the pool can be identified by measuring the gRNA populations at defined time points. In this work, four specific populations were collected and sequenced to determine the gRNA populations: the plasmid library used to generate the virus, ESCs following 12 days of selection, EpiLCs following 42 hours of differentiation, and d6 PGCLCs sorted via FACS. This allowed me to track the normalized gRNA counts as the cell developed from ESCs towards PGCLC fate, and thus identify positively or negatively enriched gRNAs between defined cell states.

The screen was designed to allow for the unbiased identification of CREs, as the *Tfap2c* TAD was completely tiled without consideration for known features. This differs from published work wherein known enhancer elements [152; 305; 306], 100 kilobase (kb) regions surrounding specific genes [157], or known TF-binding motifs within CREs [307; 308] were tiled with gRNAs. As enhancers may act on distant genes [309] and most interactions occur between CREs and genes located within the same TAD [160], the entire *Tfap2c* TAD was tiled to identify any possible CREs acting on the gene. Additionally, the screen was designed to track gRNA enrichment over developmental time points, which has not previously been shown in a CRE-targeting screen - other works cited here primarily examined one cell type and one selection method, *e.g.* Sanjana *et al.* examined mutations affecting cancer cell resistance to drug treatment [157], while Canver *et al.* examined mutations affecting *Pou5f1* expression in ESCs [158].

#### 4.2.2 gRNA enrichment is demonstrated in each examined cell type

The screen was employed in total three times, and gDNA isolated from ESCs, EpiLCs, and d6 GFP+/tdTomato- PGCLCs collected during the screen was used to generate amplicon libraries which were subsequently sequenced. 328 out of 29,192 (1.12%) unique test gRNAs, meaning gRNAs which were not taken from the GeCKO library as exonic or scrambled controls, were considered significant. Of the total gRNAs, 129 were identified in ESCs, of which 105 were negatively and 24 were positively enriched, 199 were identified in EpiLCs, of which 147 were negative and 52 were positive, and 53 were identified in d6 GFP+/tdTomato- PGCLCs, of which 18 were negative and 35 were positive (Figure 37). Negative enrichment means that a gRNA-carrying cell is depleted from the pool of cells, indicating that the cell may have undergone apoptosis, experienced reduced proliferation, or in the case of PGCLC negative enrichment, failed to express Stella and/or GFP. Positive enrichment means that a gRNA-carrying cell is enriched in the pool of cells, indicating increased proliferation or better survival relative to cells expressing other gRNAs. Addi-



tionally, in the PGCLC dataset positive enrichment may indicate a gRNA-carrying cell was enriched in the sorted population, *i.e.* KO of *Tdtomato* should result in the enrichment of cells carrying a *Tdtomato*-targeting gRNAs, as it moves the population from the unselected tdTomato+ population potentially into the sorted GFP+/tdTomato- population. It was noted that most significant gRNAs in the screen were identified in the EpiLC samples; it is therefore possible that the *Tfap2c* TAD harbors CREs required for or enhancing the transition to formative pluripotency. Further, most significant gRNAs identified in the ESCs and EpiLCs were negatively enriched, while in the PGCLCs most gRNAs were positively enriched. This indicates that the gRNAs employed in this screen primarily targeted genomic coordinates associated with reduced cellular fitness in the ESCs and EpiLCs, while coordinates targeted in the PGCLCs may have enhanced either proliferative ability or induction efficiency towards PGCLC fate.

#### 4.2.2.1 gRNAs primarily target non-coding regions in ESCs

Control gRNAs were employed during this screen to target PGCLC marker genes or genes involved in the sorting of PGCLCs at the FACS, including *Prdm1*, *Prdm14*, *Dppa3*, *Dppa5a*, *Gfp*, *Tfap2c*, and *Tdtomato*. These genes were targeted as they are either upregulated in PGCs and thus should appear negatively enriched in the PGCLC samples, or they target genes which otherwise would exclude samples from being sorted for collection at the FACS, *i.e.* *Dppa5a* and *Tdtomato*, and thus should appear positively enriched in the PGC samples. In principle, these PGC-specific controls should allow for both cell type and fluorescence-based evaluation of screen efficacy. Although they were designed to control the PGCLC population, enrichment identified in the ESCs may nevertheless be discussed. In the ESCs, gRNAs targeting *Dppa3* and *Dppa5a* were found to be significantly negatively enriched, while gRNAs targeting *Prdm14* were found to be positively enriched (Table 2). It is known that both *Dppa3* and *Dppa5a* are expressed in ESCs [11], and KO of *Dppa3* is generally embryonic lethal due to cleavage defects in early embryogenesis [310], so it is unsurprising that ESCs expressing *Dppa3*-targeting gRNAs would experience impaired cellular fitness. However KO of *Dppa5a* is non-lethal in the mouse [311], and thus the negative enrichment was unanticipated. One possible explanation for the negative enrichment of *Dppa5a*-targeting gRNAs is the fact that, although the sequences were taken directly from the GeCKO library, I found that 5/6 of the gRNAs target multiple 100% off-target regions in the genome when examined using the UCSC BLAT (BLAST-like alignment tool) search. Thus, their enrichment may not be a reliable indicator of *Dppa5a* function in the ESCs. Further, while the KO of *Prdm14* in ESCs has been shown to not affect the survival of the cells, it did result in reduced proliferation [312].

One potential explanation for the positive enrichment of *Prdm14* is the overall effect of lentiviral transduction on ESCs. The endogenous ESC anti-viral defense response has been characterized in the context of lentiviral transduction by Geis *et al.*, and may result in silencing of transgenes [313]. Further, it has been shown that small amounts of plasmid may carry over from the lentiviral preparation into the transduction event, affecting the transduction of infected cells [314]. Finally, the small volume of lentiviral supernatant used to transduce the cells contained 10% fetal bovine serum (FBS), which affects the epigenetics and developmental potential of ESCs [315]. The cumulative effect of the silencing, plasmid carryover, and serum may affect the initial gRNA representation following transduction. Essentially, the gRNA population on day one of puromycin selection may demonstrate small differences from the gRNA population within the library, which could result in the appearance of positive or negative enrichment when the ESCs are collected after twelve days and the resulting gRNA population is compared to that of the plasmid library. Potential means of improving this process may be found in Section 4.2.3.2.

Fourteen gRNAs were identified which are negatively enriched in ESCs but positively enriched in EpiLCs. Out these fourteen, two were found to be correspondingly negatively enriched in PGCLCs. It is possible that these gRNAs are targeting TF binding sites associated with naïve pluripotency, and should be examined closely. In total, eight negatively enriched and five positively enriched gRNAs found to be significant in ESCs were located in ENCODE candidate CREs (cCREs) (Figure 39), and these gRNAs should also be carefully considered for future validation. Nineteen gRNAs targeting potential regulatory region 1 (PRR1; chr2:170,246,202-170,503,872) were identified to be significant in the negatively enriched ESC dataset. PRR1 contains putative regulatory regions identified through Hi-C chromatin capture to be interacting with the *Tfap2c* promoter in ESCs and EpiLCs. Of these gRNAs, two were also located in cCREs and five were located within the intron of the gene *Bcas1* (Brain enriched myelin associated protein 1), a protein associated with myelination [316]. The numerous gRNAs identified in PRR1 indicate there may be a functionally relevant role for this region in ESC regulation, and should be investigated further.

It should be noted that gRNAs targeting control enhancers identified to be necessary for maintenance of the ESC state [267], namely the positive control enhancers (PCEs) 1-4 (Appendix 7.11), were largely not detected in the negatively enriched gRNA datasets, with the exception of one gRNA detected in PCE3 (chr14\_A\_12, targeting chr14:99,324,398-99,324,417). The control enhancers identified in the Kearns *et al.* paper were found by targeting them with Cas9-LSD1 (Cas9 fused to lysine-specific demethylase 1A), a transcriptional repressor, and looking for changes in morphology and expression of naïve pluripotency genes [267]. As such, single TF binding sites were not characterized or mapped, rather the

repression of the enhancer as a whole was used to determine its effect upon ESC state. Thus, the failure to detect enriched gRNAs in these enhancers may not be a failure of the screen, but rather it may be that few or no individual sites within these enhancers are sufficient to impair ESC state maintenance.

#### 4.2.2.2 Strong negative enrichment of gRNAs targeting exons in EpiLCs

Once the gRNA enrichment of ESCs had been determined, gRNA enrichment in EpiLCs was examined. It should be noted that gRNA enrichment analysis in EpiLCs was found by comparing normalized EpiLC gRNA populations to the normalized ESC gRNA population, and not the plasmid library. This represents an improved method of examination compared to the method employed in the ESCs, as any changes detected in the normalized gRNA population of the EpiLCs may be directly compared to the normalized gRNA population existent at the moment of EpiLC induction in the ESCs. Analysis of PGC gene-targeting control gRNAs determined to be significant in the EpiLCs found that again *Prdm14*-targeting gRNAs were positively enriched, while *Dppa3*-targeting gRNAs were negatively enriched (Table 2). As *Prdm14* is a naïve pluripotency factor which is downregulated in the EpiLCs (Figure 12), it is plausible that its KO may in fact enhance the proliferation of the formatively pluripotent EpiLCs. The negative depletion of *Dppa3*-targeting gRNAs may be explained by the fact that *Dppa3*-KO has been shown to affect global DNA methylation and misregulation of imprinting [275], and thus it is plausible that KO of the gene may in fact impair EpiLC induction and subsequent proliferation, if the ESC state is already impaired by misregulated gene expression.

53% of all significant negatively enriched gRNAs in the EpiLCs targeted exons, primarily belonging to the genes *Aurka*, *Cstf1*, and *Rae1*. AURKA-KO ESCs demonstrate reduced proliferation, enhanced differentiation, and cell death [317]. CSTF1 contributes to a complex involved in the processing of mRNA poly(A) tails, and its KO results in reduced affinity of the complex to the mRNA which strongly suggests a fundamental biological role [269]. RAE1-KO results in embryonic lethality in the mouse, is required for segregation of chromosomes, and is involved in regulation of the mitotic checkpoint [270]. These genes are required for survival and exert extremely important cellular functions, so the single gRNA readout of individual exons strongly indicates that the screen was successful at identifying single gRNAs resulting in reduced cellular fitness, and that the CRISPR/Cas9 approach was successfully employed. As the gRNAs targeting these genes were not detected to be significant in the ESCs, it is possible that they are in fact required for EpiLC induction or maintenance.

gRNAs targeting cCREs were only found to be significant in the EpiLCs within the nega-

tively enriched dataset. Of these, eight are unique to EpiLCs and are not detected in other cell types, and should be closely examined as potentially targeting regulatory sites involved in formative pluripotency. The remaining four gRNAs are also identified in the positively enriched PGCLC datasets. It is possible that these represent TF binding sites which impair the formative pluripotent state and enhance the proliferation or induction of PGCLCs, and should also be closely considered. Eight gRNAs were found to be negatively enriched in PRR1, and may also be considered for further exploration.

#### 4.2.2.3 gRNA enrichment in PGCLCs is primarily positive

Finally, the significant gRNA enrichment of PGCLCs was assessed, as the final stage of the screen and the only stage requiring reporter gene-based FACS sorting. gRNA enrichment in PGCLCs was determined by comparing the normalized PGCLC gRNA population against the EpiLC population, and not the gRNA library. Control gene-targeting gRNA enrichment, when detected, demonstrates that most controls were successfully detected within the expected positive or negative enrichment dataset (Table 2): gRNAs targeting PGC marker genes *Dppa3*, *Prdm14* and *Dppa3*-reporting *Gfp* were found to be negatively enriched, while *TdTomato*-targeting gRNAs were found to be positively enriched, as expected. Notably, gRNAs targeting *Dppa5a* directly were not found to be significant. However, it should again be noted that the *Dppa5a* gRNAs, although taken directly from the published GeCKO library, demonstrate numerous off-targets. For this reason, the *Dppa5a*-targeting gRNAs must be assessed critically.

gRNAs targeting *Tfap2c* and *Prdm1* were not detected. One possible explanation for the failure to detect these key gRNAs targeting PGCLCs genes, and specifically the gene of interest for this screen *Tfap2c*, may be low overall coverage, *i.e.* the number of total PGCLCs collected fell below the threshold of 500 cells/gRNA suggested by Joung *et al.* [156] or 300 cells/gRNA suggested by Dr. Ophir Shalem and Dr. Neville Sanjana, noted researchers in the field of screen-based technology [318]. It is challenging to collect sufficient PGCLCs, as PGCLC generation within EBs is not 100% efficient, as shown in Figure 22 representing the average percentage of putative PGCLCs isolated from d6 SGET +cyto EBs. Additionally, the dissociation and processing of EBs is also a technical, time-consuming process which limits the number of PGCLCs which can realistically be isolated on one day. For this reason, I utilized microwell plates to generate the PGCLCs and increase the overall number of cells which could be harvested upon d6 of PGCLC induction. Ultimately, I was able to achieve an average coverage of ~128 PGCLCs/gRNA. This may explain why the screen failed to detect gRNAs targeting either *Tfap2c* or *Prdm1*, in spite of the fact that *Tfap2c*-KO results in a significant decrease in GFP+/tdTomato- cells in S-KO d6 +cyto

EBs relative to WT (Figure 22). It can thus be assumed that the gRNAs determined to be significant in the d6 PGCLCs are not exclusively targeting CREs regulating *Tfap2c*.

Nevertheless, interesting gRNAs were identified, and their target regions may be further investigated for their role in overall PGCLC induction. These include three gRNAs targeting CREs in the negatively enriched gRNA dataset and a gRNA identified to be negatively depleted in the ESCs as well as the PGCLCs (chr2C\_2313, chr2:172,128,088-172,128,107). A greater number of significant gRNAs were identified as positively enriched than negatively enriched in the PGCLCs, which is intriguing. Of these, three were found to be located within cCREs. It is possible that these gRNAs target TF binding sites which may negatively regulate *Dppa3* expression or PGCLC induction, and as such should be investigated further. Four gRNAs in PRR1 were negatively enriched, of which two were located in a *Bcas1* intron, and these may also represent interesting regulatory sites within a region predicted to regulate the *Tfap2c* promoter. It should be noted that significant gRNAs may be targeting CREs which act on *Tfap2c* in addition to exerting cell type-specific activity. Thus, the possibility of identifying CREs acting on *Tfap2c* using this data remains open.

### 4.2.3 Conclusions on *Tfap2c* tiling screen

It may be concluded that the *Tfap2c* tiling screen described in this work successfully identified most control gRNAs on d6 of PGCLC induction and successfully identified on a single-gRNA resolution individual exons of putative housekeeping genes in the EpiLCs. Additionally, numerous gRNAs were localized to cCREs as well as introns, which are known to harbor CREs [135], indicating that the gRNAs may be detecting TF binding sites within putative CREs. Methods to identify single gRNAs for further examination are discussed in this conclusion, as well as potential methods to improve tiling screens in the future.

#### 4.2.3.1 Identification of single gRNAs for further characterization

With the significant gRNAs presented in this work, it is possible to identify particular gRNAs which should be investigated further to determine whether the gRNA truly exerts an influence on the corresponding cell type. Examination of individual gRNAs could begin in ESCs, as the effect upon the naïve pluripotent state including differentiation or reduced proliferation is easy to observe. Particular, significant gRNAs of interest whose target sequence could be examined for function in ESCs include: chr2C\_6355 (chr2:172,440,868-172,440,887) which is located within a cCRE negatively enriched in both ESCs and EpiLCs; chr2C\_18114 (chr2:173,171,662-173,171,681), which is located within a cCRE and is negatively enriched in ESCs and positively enriched in EpiLCs; or chr2B\_1266 (chr2:170,343,597-170,343,616), which is located within a cCRE in PRR1.

To improve the identification of promising regulatory elements within the datasets, significant gRNAs may be compared to the ATAC-seq data presented in this work as well as published H3K4Me1/H3K27ac ChIP-seq datasets to characterize enhancer-related histone modifications and chromatin accessibility. Subsequently, motif enrichment analysis may be performed upon these gRNAs to identify if specific TF binding motifs are present within these regions, and the expression of identified TFs in each cell type may be confirmed via the RNA-seq data presented in this work. Sites targeted by individual gRNAs may be functionally investigated by using the gRNA in CRISPR/Cas9 to repeat the targeted mutagenesis in a single line, and subsequently characterize the resulting phenotype. Further, inducible dCas9-KRAB (Nuclease-dead Cas9 Krüppel-associated box) could be targeted to CREs of interest to repress enhancer activity at different points of development, and the resulting phenotype may be characterized. Finally, luciferase assays could be employed to validate the interaction of the CRE of interest with a promoter element of a gene identified to be of interest following CRE characterization [319].

#### 4.2.3.2 Improvements suggested for further screens

To improve future screens, and specifically to control for potential variations in gRNA representation within cells following transduction, a Dox-inducible Cas9-expressing ESC line could be employed. In this model, a cell line is stably transfected with the reverse tetracycline-controlled transactivator (rtTA) transgene as well as Cas9 under the control of the tet response element (TRE) [320]. These cells will thus express Cas9 upon the addition of tetracycline or doxycycline (dox). Thus, these cells could be transduced with lentivirus to allow stable integration of the gRNA, passaged to ensure complete reacclimatization to the 2i+LIF medium and ensure serum-specific effects are lost, and a pellet of the cells may be collected for gDNA isolation and sequencing, with this sample representing the day 0/control sample for the ESCs instead of the plasmid library. The cells could then be treated with Dox to induce Cas9 expression, grown for twelve days, and collected. The sequenced gRNA populations in these cells could then be directly compared to the ESCs prior to the addition of Dox and thus Cas9 expression, and may provide an improved method to more sensitively determine changes in gRNA enrichment following CRISPR/Cas9 mediated mutagenesis.

Future screens could employ a protospacer adjacent motif (PAM)-free or reduced-PAM Cas9. The *S. pyogenes* Cas9 enzyme generally employed in CRISPR/Cas9 screen recognizes the PAM site 5' - NGG - 3', however reduced-PAM Cas9 variants recognize broader sites and thus dramatically increase the number of possible gRNAs which can be used to tile a region, improving overall coverage [321]. Finally, future screen could employ control regions

identified in comparable single-gRNA screens, for example the Canver *et al.* screened published in 2020 which identified on the single-gRNA level regulatory elements acting on core pluripotency factor *Pou5f1*.





# Bibliography

- [1] Melanie D. White, Jennifer Zenker, Stephanie Bissiere, and Nicolas Plachta. Instructions for Assembling the Early Mammalian Embryo. *Developmental Cell*, 45(6):667–679, 2018-06. ISSN 15345807. doi: 10.1016/j.devcel.2018.05.013.
- [2] Elisabeth Clare Larsen, Ole Bjarne Christiansen, Astrid Marie Kolte, and Nick Macklon. New insights into mechanisms behind miscarriage. *BMC Medicine*, 11(154), 2013. doi: 10.1186/1741-7015-11-154.
- [3] Melanie A. Eckersley-Maslin, Celia Alda-Catalinas, and Wolf Reik. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nature Reviews. Molecular Cell Biology*, 19(7):436–450, 2018-07. ISSN 1471-0080. doi: 10.1038/s41580-018-0008-z.
- [4] Adam Burton and Maria-Elena Torres-Padilla. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nature Reviews Molecular Cell Biology*, 15(11):723–735, 2014-11. ISSN 1471-0080. doi: 10.1038/nrm3885.
- [5] Adam Burton, Julius Muller, Shengjiang Tu, Pablo Padilla-Longoria, Ernesto Guccione, and Maria-Elena Torres-Padilla. Single-cell profiling of epigenetic modifiers identifies PRDM14 as an inducer of cell fate in the mammalian embryo. *Cell Reports*, 5(3):687–701, 2013-11-14. ISSN 2211-1247. doi: 10.1016/j.celrep.2013.09.044.
- [6] Natalia Gawlik-Rzemieniewska and Ilona Bednarek. The role of NANOG transcriptional factor in the development of malignant phenotype of cancer cells. *Cancer Biology & Therapy*, 17(1):1–10, 2015-11-30. ISSN 1538-4047. doi: 10.1080/15384047.2015.1121348.
- [7] Yusuke Miyanari and Maria-Elena Torres-Padilla. Control of ground-state pluripotency by allelic regulation of Nanog. *Nature*, 483(7390):470–473, 2012-03. ISSN 1476-4687. doi: 10.1038/nature10807.
- [8] Berenika Plusa and Anna Piliszek. Common principles of early mammalian embryo self-organisation. *Development*, 147(14), 2020. doi: 10.1242/dev.183079.
- [9] Ecem Yildirim, Gizem Bora, Tugce Onel, Nilsu Talas, and Aylin Yaba. Cell fate determination and Hippo signaling pathway in preimplantation mouse embryo. *Cell and Tissue Research*, 386(3):423–444, 2021-12-01. ISSN 1432-0878. doi: 10.1007/s00441-021-03530-8.
- [10] Janet Rossant. Genetic Control of Early Cell Lineages in the Mammalian Embryo. *Annual Review of Genetics*, 52:185–201, 2018-09-05. doi: 10.1146/annurev-genet-120116-024544.
- [11] Jamie A. Hackett and M. Azim Surani. Regulatory Principles of Pluripotency: From the Ground State Up. *Cell Stem Cell*, 15(4):416–430, 2014-10-02. ISSN 1934-5909. doi: 10.1016/j.stem.2014.09.015.
- [12] Daniela Ávila-González, Wendy Portillo, Guadalupe García-López, Anayansi Molina-Hernández, Néstor E. Díaz-Martínez, and Néstor F. Díaz. Unraveling the Spatiotemporal Human Pluripotency in Embryonic Development. *Frontiers in Cell and Developmental Biology*, 9, 2021. ISSN 2296-634X. URL <https://www.frontiersin.org/article/10.3389/fcell.2021.676998>.
- [13] Yang Zeng and Taiping Chen. DNA Methylation Reprogramming during Mammalian Development. *Genes*, 10(4):E257, 2019-03-29. ISSN 2073-4425. doi: 10.3390/genes10040257.
- [14] Jose Silva, Ornella Barrandon, Jennifer Nichols, Jitsutaro Kawaguchi, Thorold W. Theunissen, and Austin Smith. Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition. *PLoS Biology*, 6(10):e253, 2008-10. ISSN 1545-7885. doi: 10.1371/journal.pbio.0060253.
- [15] Marta N. Shahbazi and Magdalena Zernicka-Goetz. Deconstructing and reconstructing the mouse and human early embryo. *Nature Cell Biology*, 20(8):878–887, 2018-08. ISSN 1476-4679. doi: 10.1038/s41556-018-0144-x.

- [16] Hiromichi Matsumoto. Molecular and cellular events during blastocyst implantation in the receptive uterus: Clues from mouse models. *The Journal of Reproduction and Development*, 63(5):445–454, 2017-10. ISSN 0916-8818. doi: 10.1262/jrd.2017-047.
- [17] Jaime A. Rivera-Pérez and Anna-Katerina Hadjantonakis. The Dynamics of Morphogenesis in the Early Mouse Embryo. *Cold Spring Harbor Perspectives in Biology*, 7(11):a015867, 2015-11. ISSN 1943-0264. doi: 10.1101/cshperspect.a015867.
- [18] Dong Hu and James C. Cross. Development and function of trophoblast giant cells in the rodent placenta. *International Journal of Developmental Biology*, 54(2-3):341–354, 2009-10-02. ISSN 0214-6282, 1696-3547. doi: 10.1387/ijdb.082768dh.
- [19] Grace V. Hancock, Sissy E. Wamaitha, Lior Peretz, and Amander T. Clark. Mammalian primordial germ cell specification. *Development (Cambridge, England)*, 148(6):dev189217, 2021-03-15. ISSN 0950-1991. doi: 10.1242/dev.189217.
- [20] Masaki Kinoshita, Michael Barber, William Mansfield, Yingzhi Cui, Daniel Spindlow, Giuliano Giuseppe Stirparo, Sabine Dietmann, Jennifer Nichols, and Austin Smith. Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. *Cell Stem Cell*, 28(3):453–471.e8, 2021-03. ISSN 19345909. doi: 10.1016/j.stem.2020.11.005.
- [21] Katsuhiko Hayashi, Hiroshi Ohta, Kazuki Kurimoto, Shinya Aramaki, and Mitinori Saitou. Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells. *Cell*, 146(4):519–532, 2011-08. ISSN 00928674. doi: 10.1016/j.cell.2011.06.052.
- [22] Alexandra Schauer and Carl-Philipp Heisenberg. Reassembling gastrulation - ScienceDirect. *Developmental Biology*, 474:71–81, 2021-06. doi: 10.1016/j.ydbio.2020.12.014.
- [23] Guojun Sheng, Alfonso Martinez Arias, and Ann Sutherland. The primitive streak and cellular principles of building an amniote body through gastrulation. *Science*, 374(6572), 2021-12-03. doi: 10.1126/science.abg1727.
- [24] Walfred W. C. Tang, Toshihiro Kobayashi, Naoko Irie, Sabine Dietmann, and M. Azim Surani. Specification and epigenetic programming of the human germ line. *Nature Reviews Genetics*, 17(10):585–600, 2016-10. ISSN 1471-0064. doi: 10.1038/nrg.2016.88.
- [25] M. Azim Surani. Germ cells: The eternal link between generations. *Comptes Rendus Biologies*, 330(6-7):474–478, 2007-06. ISSN 16310691. doi: 10.1016/j.crv.2007.03.009.
- [26] Yasuhide Ohinata, Bernhard Payer, Dónal O’Carroll, Katia Ancelin, Yukiko Ono, Mitsue Sano, Sheila C. Barton, Tetyana Obukhanych, Michel Nussenzweig, Alexander Tarakhovskiy, Mitinori Saitou, and M. Azim Surani. *Blimp1* is a critical determinant of the germ cell lineage in mice. *Nature*, 436(7048):207–213, 2005-07. ISSN 1476-4687. doi: 10.1038/nature03813.
- [27] Yasuhide Ohinata, Hiroshi Ohta, Mayo Shigeta, Kaori Yamanaka, Teruhiko Wakayama, and Mitinori Saitou. A signaling principle for the specification of the germ cell lineage in mice. *Cell*, 137(3):571–584, 2009-05-01. ISSN 1097-4172. doi: 10.1016/j.cell.2009.03.014.
- [28] K. A. Lawson, N. R. Dunn, B. A. Roelen, L. M. Zeinstra, A. M. Davis, C. V. Wright, J. P. Korving, and B. L. Hogan. *Bmp4* is required for the generation of primordial germ cells in the mouse embryo. *Genes & Development*, 13(4):424–436, 1999-02-15. ISSN 0890-9369. doi: 10.1101/gad.13.4.424.
- [29] Y. Ying and G. Q. Zhao. Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. *Developmental Biology*, 232(2):484–492, 2001-04-15. ISSN 0012-1606. doi: 10.1006/dbio.2001.0173.
- [30] Y. Ying, X. M. Liu, A. Marble, K. A. Lawson, and G. Q. Zhao. Requirement of *Bmp8b* for the generation of primordial germ cells in the mouse. *Molecular Endocrinology (Baltimore, Md.)*, 14(7):1053–1063, 2000-07. ISSN 0888-8809. doi: 10.1210/mend.14.7.0479.

- [31] Yigong Shi and Joan Massagué. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell*, 113(6):685–700, 2003-06-13. ISSN 0092-8674. doi: 10.1016/s0092-8674(03)00432-x.
- [32] Joan Massagué, Joan Seoane, and David Wotton. Smad transcription factors. *Genes & Development*, 19(23):2783–2810, 2005-12-01. ISSN 0890-9369. doi: 10.1101/gad.1350705.
- [33] P. Liu, M. Wakamiya, M. J. Shea, U. Albrecht, R. R. Behringer, and A. Bradley. Requirement for Wnt3 in vertebrate axis formation. *Nature Genetics*, 22(4):361–365, 1999-08. ISSN 1061-4036. doi: 10.1038/11932.
- [34] Shinya Aramaki, Katsuhiko Hayashi, Kazuki Kurimoto, Hiroshi Ohta, Yukihiro Yabuta, Hiroko Iwanari, Yasuhiro Mochizuki, Takao Hamakubo, Yuki Kato, Katsuhiko Shirahige, and Mitinori Saitou. A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Developmental Cell*, 27(5):516–529, 2013-12-09. ISSN 1878-1551. doi: 10.1016/j.devcel.2013.11.001.
- [35] Susanne Weber, Dawid Eckert, Daniel Nettersheim, Ad J. M. Gillis, Sabine Schäfer, Peter Kuckenberger, Julia Ehlermann, Uwe Werling, Katharina Biermann, Leendert H. J. Looijenga, and Hubert Schorle. Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. *Biology of Reproduction*, 82(1):214–223, 2010-01. ISSN 1529-7268. doi: 10.1095/biolreprod.109.078717.
- [36] Masashi Yamaji, Yoshiyuki Seki, Kazuki Kurimoto, Yukihiro Yabuta, Mihoko Yuasa, Mayo Shigeta, Kaori Yamanaka, Yasuhide Ohinata, and Mitinori Saitou. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nature Genetics*, 40(8):1016–1022, 2008-08. ISSN 1546-1718. doi: 10.1038/ng.186.
- [37] Erna Magnúsdóttir, Sabine Dietmann, Kazuhiro Murakami, Ufuk Günesdogan, Fuchou Tang, Siqin Bao, Evangelia Diamanti, Kaiqin Lao, Berthold Gottgens, and M. Azim Surani. A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nature Cell Biology*, 15(8):905–915, 2013-08. ISSN 1476-4679. doi: 10.1038/ncb2798.
- [38] Erna Magnúsdóttir and M. Azim Surani. How to make a primordial germ cell. *Development (Cambridge, England)*, 141(2):245–252, 2014-01. ISSN 1477-9129. doi: 10.1242/dev.098269.
- [39] Mitinori Saitou, Sheila C. Barton, and M. Azim Surani. A molecular programme for the specification of germ cell fate in mice. *Nature*, 418(6895):293–300, 2002-07-18. ISSN 0028-0836. doi: 10.1038/nature00927.
- [40] Kazuki Kurimoto, Yukihiro Yabuta, Yasuhide Ohinata, Mayo Shigeta, Kaori Yamanaka, and Mitinori Saitou. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes & Development*, 22(12):1617–1635, 2008-06-15. ISSN 0890-9369. doi: 10.1101/gad.1649908.
- [41] Satomi S. Tanaka and Ryuichi Nishinakamura. Regulation of male sex determination: Genital ridge formation and Sry activation in mice. *Cellular and Molecular Life Sciences*, 71(24):4781–4802, 2014. ISSN 1420-682X. doi: 10.1007/s00018-014-1703-3.
- [42] Daniel M. Messerschmidt, Barbara B. Knowles, and Davor Solter. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes & Development*, 28(8):812–828, 2014-04-15. ISSN 1549-5477. doi: 10.1101/gad.234294.113.
- [43] Kathleen A. Molyneaux, Jim Stallock, Kyle Schaible, and Christopher Wylie. Time-Lapse Analysis of Living Mouse Germ Cell Migration. *Developmental Biology*, 240(2):488–498, 2001-12-15. ISSN 0012-1606. doi: 10.1006/dbio.2001.0436.
- [44] Yoshiyuki Seki, Katsuhiko Hayashi, Kunihiko Itoh, Michinao Mizugaki, Mitinori Saitou, and Yasuhisa Matsui. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Developmental Biology*, 278(2):440–458, 2005-02-15. ISSN 0012-1606. doi: 10.1016/j.ydbio.2004.11.025.
- [45] Stefanie Seisenberger, Simon Andrews, Felix Krueger, Julia Arand, Jörn Walter, Fátima Santos, Christian Popp, Bernard Thienpont, Wendy Dean, and Wolf Reik. The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Molecular Cell*, 48(6):849–862, 2012-12-28. ISSN 1097-2765. doi: 10.1016/j.molcel.2012.11.001.

- [46] Katia Ancelin, Ulrike C. Lange, Petra Hajkova, Robert Schneider, Andrew J. Bannister, Tony Kouzarides, and M. Azim Surani. Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nature Cell Biology*, 8(6):623–630, 2006-06. ISSN 1465-7392. doi: 10.1038/ncb1413.
- [47] Hongshan Guo, Boqiang Hu, Liying Yan, Jun Yong, Yan Wu, Yun Gao, Fan Guo, Yu Hou, Xiaoying Fan, Ji Dong, Xiaoye Wang, Xiaohui Zhu, Jie Yan, Yuan Wei, Hongyan Jin, Wenxin Zhang, Lu Wen, Fuchou Tang, and Jie Qiao. DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. *Cell Research*, 27(2):165–183, 2017-02. ISSN 1748-7838. doi: 10.1038/cr.2016.128.
- [48] Michihiko Sugimoto and Kuniya Abe. X chromosome reactivation initiates in nascent primordial germ cells in mice. *PLoS genetics*, 3(7):e116, 2007-07. ISSN 1553-7404. doi: 10.1371/journal.pgen.0030116.
- [49] Jillian Guttormsen, prefix=de la useprefix=true family=Houssaye, given=Christopher, Anne Howlett, Frances Bhushan, Trevor Williams, and Quinton Winger. DISRUPTION OF GERM CELL DEVELOPMENT IN TRANSCRIPTION FACTOR AP-2GAMMA MUTANT MICE. *Biology of Reproduction*, 77:197, 2007-07-01. ISSN 0006-3363. doi: 10.1093/biolreprod/77.s1.197a.
- [50] Jana Schemmer, Marcos J. Araúzo-Bravo, Natalie Haas, Sabine Schäfer, Susanne N. Weber, Astrid Becker, Dawid Eckert, Andreas Zimmer, Daniel Nettersheim, and Hubert Schorle. Transcription factor TFAP2C regulates major programs required for murine fetal germ cell maintenance and haploinsufficiency predisposes to teratomas in male mice. *PLoS One*, 8(8):e71113, 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0071113.
- [51] Katharina Pauls, Hubert Schorle, Wiebke Jeske, Ralph Brehm, Klaus Steger, Nicolas Wernert, Reinhard Büttner, and Hui Zhou. Spatial expression of germ cell markers during maturation of human fetal male gonads: An immunohistochemical study. *Human Reproduction*, 21(2):397–404, 2006-02. ISSN 1460-2350. doi: 10.1093/humrep/dei325.
- [52] Di Chen, Wanlu Liu, Jill Zimmerman, William A. Pastor, Rachel Kim, Linzi Hosohama, Jamie Ho, Marianna Aslanyan, Joanna J. Gell, Steven E. Jacobsen, and Amander T. Clark. The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation. *Cell Reports*, 25(13):3591–3602.e5, 2018-12-26. ISSN 2211-1247. doi: 10.1016/j.celrep.2018.12.011.
- [53] Toshihiro Kobayashi, Aracely Castillo-Venzor, Chris A. Penfold, Michael Morgan, Naoaki Mizuno, Walfred W. C. Tang, Yasuyuki Osada, Masao Hirao, Fumika Yoshida, Hideyuki Sato, Hiromitsu Nakauchi, Masumi Hirabayashi, and M. Azim Surani. Tracing the emergence of primordial germ cells from bilaminar disc rabbit embryos and pluripotent stem cells. *Cell Reports*, 37(2):109812, 2021-10-12. ISSN 2211-1247. doi: 10.1016/j.celrep.2021.109812.
- [54] prefix=de useprefix=true family=Souza, given=Aline Fernanda, Naira Caroline Godoy Pieri, and Daniele dos Santos Martins. Step by Step about Germ Cells Development in Canine. *Animals : an Open Access Journal from MDPI*, 11(3):598, 2021-02-25. ISSN 2076-2615. doi: 10.3390/ani11030598.
- [55] Qifan Zhu, Fei Sang, Sarah Withey, Walfred Tang, Sabine Dietmann, Doris Klisch, Priscila Ramos-Ibeas, Haixin Zhang, Cristina E. Requena, Petra Hajkova, Matt Loose, M. Azim Surani, and Ramiro Alberio. Specification and epigenomic resetting of the pig germline exhibit conservation with the human lineage. *Cell Reports*, 34(6):108735, 2021-02-09. ISSN 2211-1247. doi: 10.1016/j.celrep.2021.108735.
- [56] Kotaro Sasaki, Tomonori Nakamura, Ikuhiro Okamoto, Yukihiko Yabuta, Chizuru Iwatani, Hideaki Tsuchiya, Yasunari Seita, Shinichiro Nakamura, Naoto Shiraki, Tetsuya Takakuwa, Takuya Yamamoto, and Mitinori Saitou. The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. *Developmental Cell*, 39(2):169–185, 2016-10-24. ISSN 1878-1551. doi: 10.1016/j.devcel.2016.09.007.
- [57] Delia Alba Soto and Pablo Juan Ross. Similarities between bovine and human germline development revealed by single-cell RNA sequencing. *Reproduction (Cambridge, England)*, 161(3):239–253, 2021-03. ISSN 1741-7899. doi: 10.1530/REP-20-0313.
- [58] C. A. Turner, D. H. Mack, and M. M. Davis. Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell*, 77(2):297–306, 1994-04-22. ISSN 0092-8674. doi: 10.1016/0092-8674(94)90321-2.

- [59] I-Ying Lin, Feng-Lan Chiu, Chen-Hsiang Yeang, Hsin-Fu Chen, Ching-Yu Chuang, Shii-Yi Yang, Pei-Shan Hou, Nardnisa Sintupisut, Hong-Nerng Ho, Hung-Chih Kuo, and Kuo-I Lin. Suppression of the SOX2 Neural Effector Gene by PRDM1 Promotes Human Germ Cell Fate in Embryonic Stem Cells. *Stem Cell Reports*, 2(2):189–204, 2014-01-23. ISSN 2213-6711. doi: 10.1016/j.stemcr.2013.12.009.
- [60] Anastasiya Sybirna, Walfred W. C. Tang, Merrick Pierson Smela, Sabine Dietmann, Wolfram H. Gruhn, Ran Brosh, and M. Azim Surani. A critical role of PRDM14 in human primordial germ cell fate revealed by inducible degrens. *Nature Communications*, 11(1):1282, 2020-03-09. ISSN 2041-1723. doi: 10.1038/s41467-020-15042-0.
- [61] Yoshiyuki Seki. PRDM14 Is a Unique Epigenetic Regulator Stabilizing Transcriptional Networks for Pluripotency. *Frontiers in Cell and Developmental Biology*, 6:12, 2018-02-13. ISSN 2296-634X. doi: 10.3389/fcell.2018.00012.
- [62] Ziyang Ma, Tomek Swigut, Anton Valouev, Alvaro Rada-Iglesias, and Joanna Wysocka. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nature Structural & Molecular Biology*, 18(2):120–127, 2011-02. ISSN 1545-9985. doi: 10.1038/nsmb.2000.
- [63] Nils Grabole, Julia Tischler, Jamie A Hackett, Shinseog Kim, Fuchou Tang, Harry G Leitch, Erna Magnúsdóttir, and M Azim Surani. Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation. *EMBO Reports*, 14(7):629–637, 2013-07. ISSN 1469-221X. doi: 10.1038/embor.2013.67.
- [64] Kenjiro Shirane, Kazuki Kurimoto, Yukihiro Yabuta, Masashi Yamaji, Junko Satoh, Shinji Ito, Akira Watanabe, Katsuhiko Hayashi, Mitinori Saitou, and Hiroyuki Sasaki. Global Landscape and Regulatory Principles of DNA Methylation Reprogramming for Germ Cell Specification by Mouse Pluripotent Stem Cells. *Developmental Cell*, 39(1):87–103, 2016-10-10. ISSN 1534-5807. doi: 10.1016/j.devcel.2016.08.008.
- [65] Harry G. Leitch and Austin Smith. The mammalian germline as a pluripotency cycle. *Development (Cambridge, England)*, 140(12):2495–2501, 2013-06. ISSN 1477-9129. doi: 10.1242/dev.091603.
- [66] Katsuhiko Hayashi, Sugako Ogushi, Kazuki Kurimoto, So Shimamoto, Hiroshi Ohta, and Mitinori Saitou. Offspring from oocytes derived from in vitro primordial germ cell-like cells in mice. *Science (New York, N. Y.)*, 338(6109):971–975, 2012-11-16. ISSN 1095-9203. doi: 10.1126/science.1226889.
- [67] Fumio Nakaki, Katsuhiko Hayashi, Hiroshi Ohta, Kazuki Kurimoto, Yukihiro Yabuta, and Mitinori Saitou. Induction of mouse germ-cell fate by transcription factors in vitro. *Nature*, 501(7466):222–226, 2013-09. ISSN 1476-4687. doi: 10.1038/nature12417.
- [68] Kazuhiro Murakami, Ufuk Günesdogan, Jan J. Zyllicz, Walfred W. C. Tang, Roopsha Sengupta, Toshihiro Kobayashi, Shinseog Kim, Richard Butler, Sabine Dietmann, and M. Azim Surani. NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. *Nature*, 529(7586):403–407, 2016-01. ISSN 1476-4687. doi: 10.1038/nature16480.
- [69] Jason A. West, Srinivas R. Viswanathan, Akiko Yabuuchi, Kerianne Cunniff, Ayumu Takeuchi, In-Hyun Park, Julia E. Sero, Hao Zhu, Antonio Perez-Atayde, A. Lindsay Frazier, M. Azim Surani, and George Q. Daley. A role for Lin28 in primordial germ-cell development and germ-cell malignancy. *Nature*, 460(7257):909–913, 2009-08. ISSN 1476-4687. doi: 10.1038/nature08210.
- [70] Zubing Cao, Timothy S. Carey, Avishek Ganguly, Catherine A. Wilson, Soumen Paul, and Jason G. Knott. Transcription factor AP-2 $\gamma$  induces early Cdx2 expression and represses HIPPO signaling to specify the trophoderm lineage. *Development (Cambridge, England)*, 142(9):1606–1615, 2015-05-01. ISSN 0950-1991. doi: 10.1242/dev.120238.
- [71] M. Oulad-Abdelghani, P. Bouillet, C. Chazaud, P. Dollé, and P. Chambon. AP-2.2: A novel AP-2-related transcription factor induced by retinoic acid during differentiation of P19 embryonal carcinoma cells. *Experimental Cell Research*, 225(2):338–347, 1996-06-15. ISSN 0014-4827. doi: 10.1006/excr.1996.0184.
- [72] Jillian Guttormsen, Maranke I. Koster, John R. Stevens, Dennis R. Roop, Trevor Williams, and Quinton A. Winger. Disruption of epidermal specific gene expression and delayed skin development in AP-2 gamma mutant mice. *Developmental Biology*, 317(1):187–195, 2008-05-01. ISSN 1095-564X. doi: 10.1016/j.ydbio.2008.02.017.

- [73] Lingjie Li, Yong Wang, Jessica L. Torkelson, Gautam Shankar, Jillian M. Pattison, Hanson H. Zhen, Fengqin Fang, Zhana Duren, Jingxue Xin, Sadhana Gaddam, Sandra P. Melo, Samantha N. Piekos, Jiang Li, Eric J. Liaw, Lang Chen, Rui Li, Marius Wernig, Wing H. Wong, Howard Y. Chang, and Anthony E. Oro. TFAP2C- and p63-Dependent Networks Sequentially Rearrange Chromatin Landscapes to Drive Human Epidermal Lineage Commitment. *Cell Stem Cell*, 24(2):271–284.e8, 2019-02-07. ISSN 1875-9777. doi: 10.1016/j.stem.2018.12.012.
- [74] Erin A. Bassett, Anna Korol, Paula A. Deschamps, Reinhard Buettner, Valerie A. Wallace, Trevor Williams, and Judith A. West-Mays. Overlapping expression patterns and redundant roles for AP-2 transcription factors in the developing mammalian retina. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 241(4):814–829, 2012-04. ISSN 1097-0177. doi: 10.1002/dvdy.23762.
- [75] Vivian W. Gu, Edward Cho, Dakota T. Thompson, Victoria C. Cassady, Nicholas Borchering, Kelsey E. Koch, Vincent T. Wu, Allison W. Lorenzen, prefix=van der useprefix=true family=Heide, given=Dana M., Jeffrey R. White, Mikhail V. Kulak, Trevor Williams, Weizhou Zhang, and Ronald J. Weigel. AP-2 $\gamma$  Is Required for Maintenance of Multipotent Mammary Stem Cells. *Stem Cell Reports*, 16(1):106–119, 2021-01-12. ISSN 2213-6711. doi: 10.1016/j.stemcr.2020.12.002.
- [76] A. R. Cyr, M. V. Kulak, J. M. Park, M. V. Bogachek, P. M. Spanheimer, G. W. Woodfield, L. S. White-Baer, Y. Q. O'Malley, S. L. Sugg, A. K. Olivier, W. Zhang, F. E. Domann, and R. J. Weigel. TFAP2C governs the luminal epithelial phenotype in mammary development and carcinogenesis. *Oncogene*, 34(4):436–444, 2015-01-22. ISSN 1476-5594. doi: 10.1038/onc.2013.569.
- [77] Megan Rothstein and Marcos Simoes-Costa. Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. *Genome Research*, 30(1):35–48, 2020-01. ISSN 1088-9051. doi: 10.1101/gr.249680.119.
- [78] Antonio Mateus-Pinheiro, Nuno Dinis Alves, Nuno Sousa, and Luisa Pinto. AP2: A New Player on Adult Hippocampal Neurogenesis Regulation. *Journal of Experimental Neuroscience*, 12:117906951876689, 2018-01. ISSN 1179-0695. doi: 10.1177/1179069518766897.
- [79] Damian Kolat, Żaneta Kałuzińska, Andrzej K. Bednarek, and Elżbieta Pluciennik. The biological characteristics of transcription factors AP-2 $\alpha$  and AP-2 $\gamma$  and their importance in various types of cancers. *Bioscience Reports*, 39(3):BSR20181928, 2019-03-29. ISSN 1573-4935. doi: 10.1042/BSR20181928.
- [80] P. J. Mitchell, C. Wang, and R. Tjian. Positive and negative regulation of transcription in vitro: Enhancer-binding protein AP-2 is inhibited by SV40 T antigen. *Cell*, 50(6):847–861, 1987-09-11. ISSN 0092-8674. doi: 10.1016/0092-8674(87)90512-5.
- [81] R. B. Gaynor, C. Muchardt, Y. R. Xia, I. Klisak, T. Mohandas, R. S. Sparkes, and A. J. Lusis. Localization of the gene for the DNA-binding protein AP-2 to human chromosome 6p22.3-pter. *Genomics*, 10(4):1100–1102, 1991-08. ISSN 0888-7543. doi: 10.1016/0888-7543(91)90209-w.
- [82] M. Moser, A. Imhof, A. Pscherer, R. Bauer, W. Amselgruber, F. Sinowatz, F. Hofstädter, R. Schüle, and R. Buettner. Cloning and characterization of a second AP-2 transcription factor: AP-2 beta. *Development (Cambridge, England)*, 121(9):2779–2788, 1995-09. ISSN 0950-1991. doi: 10.1242/dev.121.9.2779.
- [83] J. A. Williamson, J. M. Boshier, A. Skinner, D. Sheer, T. Williams, and H. C. Hurst. Chromosomal mapping of the human and mouse homologues of two new members of the AP-2 family of transcription factors. *Genomics*, 35(1):262–264, 1996-07-01. ISSN 0888-7543. doi: 10.1006/geno.1996.0351.
- [84] C. Chazaud, M. Oulad-Abdelghani, P. Bouillet, D. Décimo, P. Chambon, and P. Dollé. AP-2.2, a novel gene related to AP-2, is expressed in the forebrain, limbs and face during mouse embryogenesis. *Mechanisms of Development*, 54(1):83–94, 1996-01. ISSN 0925-4773. doi: 10.1016/0925-4773(95)00463-7.
- [85] F. Zhao, M. Satoda, J. D. Licht, Y. Hayashizaki, and B. D. Gelb. Cloning and characterization of a novel mouse AP-2 transcription factor, AP-2delta, with unique DNA binding and transactivation properties. *The Journal of Biological Chemistry*, 276(44):40755–40760, 2001-11-02. ISSN 0021-9258. doi: 10.1074/jbc.M106284200.

- [86] Ramakumar Tummala, Rose-Anne Romano, Elaine Fuchs, and Satrajit Sinha. Molecular cloning and characterization of AP-2 epsilon, a fifth member of the AP-2 family. *Gene*, 321:93–102, 2003-12-04. ISSN 0378-1119. doi: 10.1016/s0378-1119(03)00840-0.
- [87] Weiguo Feng and Trevor Williams. Cloning and characterization of the mouse AP-2 epsilon gene: A novel family member expressed in the developing olfactory bulb. *Molecular and Cellular Neurosciences*, 24(2): 460–475, 2003-10. ISSN 1044-7431. doi: 10.1016/s1044-7431(03)00209-4.
- [88] Quinton Winger, Jian Huang, Heidi J. Auman, Mark Lewandoski, and Trevor Williams. Analysis of transcription factor AP-2 expression and function during mouse preimplantation development. *Biology of Reproduction*, 75(3):324–333, 2006-09. ISSN 0006-3363. doi: 10.1095/biolreprod.106.052407.
- [89] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021-08. ISSN 1476-4687. doi: 10.1038/s41586-021-03819-2.
- [90] AlphaFold. Transcription factor AP-2 gamma, 2021-07-01. URL <https://alphafold.ebi.ac.uk/entry/Q61312>.
- [91] NCBI Gene. Tfp2c transcription factor AP-2, gamma [ *Mus musculus* (house mouse) ], 2022-02-22. URL <https://www.ncbi.nlm.nih.gov/gene/21420>.
- [92] UniProt. AP2C\_MOUSE, 2022-02-23. URL <https://www.uniprot.org/uniprot/Q61312>.
- [93] UniProt. Q3ULB3, 2022-01-19. URL <https://www.uniprot.org/uniprot/Q3ULB3>.
- [94] UniProt. A2APA8, 2022-01-19. URL <https://www.uniprot.org/uniprot/A2APA8>.
- [95] John H Ludes-Meyers, Hyunsuk Kil, Andrzej K Bednarek, Jeff Drake, Mark T Bedford, and C Marcelo Aldaz. WWOX binds the specific proline-rich ligand PPXY: Identification of candidate interacting proteins. *Oncogene*, 23(29):5049–5055, 2004-06-24. ISSN 0950-9232. doi: 10.1038/sj.onc.1207680.
- [96] Karim Taouis, Keltouma Driouch, Rosette Lidereau, and François Lallemand. Molecular Functions of WWOX Potentially Involved in Cancer Development. *Cells*, 10(5):1051, 2021-04-29. ISSN 2073-4409. doi: 10.3390/cells10051051.
- [97] N. Mohibullah, A. Donner, J. A. Ippolito, and T. Williams. SELEX and missing phosphate contact analyses reveal flexibility within the AP-2[alpha] protein: DNA binding complex. *Nucleic Acids Research*, 27(13): 2760–2769, 1999-07-01. ISSN 0305-1048. doi: 10.1093/nar/27.13.2760.
- [98] George W. Woodfield, Yizhen Chen, Thomas B. Bair, Frederick E. Domann, and Ronald J. Weigel. Identification of Primary Gene Targets of TFAP2C in Hormone Responsive Breast Carcinoma Cells. *Genes, chromosomes & cancer*, 49(10):948–962, 2010-10. ISSN 1045-2257. doi: 10.1002/gcc.20807.
- [99] Zhi-Jian Han, Yan-Hu Feng, Bao-Hong Gu, Yu-Min Li, and Hao Chen. The post-translational modification, SUMOylation, and cancer (Review). *International Journal of Oncology*, 52(4):1081–1094, 2018-02-22. ISSN 1019-6439. doi: 10.3892/ijo.2018.4280.
- [100] Jyrki J. Eloranta and Helen C. Hurst. Transcription Factor AP-2 Interacts with the SUMO-conjugating Enzyme UBC9 and Is Sumolated in Vivo\*. *Journal of Biological Chemistry*, 277(34):30798–30804, 2002-08-23. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M202780200.
- [101] ENSEMBL. Gene: Tfp2a ENSMUSG00000021359 - Parologue alignment, n.d.
- [102] Colin Kenny, Ramile Dilshat, Hannah Seberg, Eric Van Otterloo, Gregory Bonde, Annika Helverson, Eiríkur Steingrímsson, and Robert A. Cornell. TFAP2 paralogs pioneer chromatin access for MITF and directly inhibit genes associated with cell migration. page 2021.11.23.469757, 2021-11-30. doi: 10.1101/2021.11.23.469757.

- [103] Päivi Pihlajamaa, Biswajyoti Sahu, Lauri Lyly, Viljami Aittomäki, Sampsa Hautaniemi, and Olli A. Jänne. Tissue-specific pioneer factors associate with androgen receptor cistromes and transcription programs. *The EMBO journal*, 33(4):312–326, 2014-02-18. ISSN 1460-2075. doi: 10.1002/embj.201385895.
- [104] Kenneth S. Zaret and Jason S. Carroll. Pioneer transcription factors: Establishing competence for gene expression. *Genes & Development*, 25(21):2227–2241, 2011-11-01. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.176826.111.
- [105] M. A. Nowak, M. C. Boerlijst, J. Cooke, and J. M. Smith. Evolution of genetic redundancy. *Nature*, 388(6638): 167–171, 1997-07-10. ISSN 0028-0836. doi: 10.1038/40618.
- [106] Dale Talbot, Janet Lorgin, and Hubert Schorle. Spatiotemporal Expression Pattern of Keratins in Skin of AP-2 $\alpha$ -deficient Mice. *Journal of Investigative Dermatology*, 113(5):816–820, 1999-11-01. ISSN 0022-202X. doi: 10.1046/j.1523-1747.1999.00759.x.
- [107] Wei Li and Robert A. Cornell. Redundant activities of Tfap2a and Tfap2c are required for neural crest induction and development of other non-neural ectoderm derivatives in zebrafish embryos. *Developmental Biology*, 304(1):338–354, 2007-04-01. ISSN 0012-1606. doi: 10.1016/j.ydbio.2006.12.042.
- [108] Hualei Li, Prabhat C Goswami, and Frederick E Domann. AP-2 $\gamma$  Induces p21 Expression, Arrests Cell Cycle, and Inhibits the Tumor Growth of Human Carcinoma Cells. *Neoplasia (New York, N.Y.)*, 8(7):568–577, 2006-07. ISSN 1522-8002. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1601932/>.
- [109] Richard Jäger, Nicolaus Friedrichs, Inge Heim, Reinhard Büttner, and Hubert Schorle. Dual role of AP-2 $\gamma$  in ErbB-2-induced mammary tumorigenesis. *Breast Cancer Research and Treatment*, 90(3):273–280, 2005-04. ISSN 0167-6806. doi: 10.1007/s10549-004-4815-x.
- [110] Ping-Pui Wong, Fabrizio Miranda, KaYi V. Chan, Chiara Berlato, Helen C. Hurst, and Angelo G. Scibetta. Histone Demethylase KDM5B Collaborates with TFAP2C and Myc To Repress the Cell Cycle Inhibitor p21cip (CDKN1A). *Molecular and Cellular Biology*, 32(9):1633–1644, 2012-05. ISSN 0270-7306. doi: 10.1128/MCB.06373-11.
- [111] Jmw Gee, Jj Eloranta, Jc Ibbitt, Jfr Robertson, Io Ellis, T Williams, Ri Nicholson, and Hc Hurst. Overexpression of TFAP2C in invasive breast cancer correlates with a poorer response to anti-hormone therapy and reduced patient survival. *The Journal of Pathology*, 217(1):32–41, 2009. ISSN 1096-9896. doi: 10.1002/path.2430.
- [112] Christina E. Hoei-Hansen, John E. Nielsen, Kristian Almstrup, Si Brask Sonne, Niels Graem, Niels E. Skakkebaek, Henrik Leffers, and Ewa Rajpert-De Meyts. Transcription factor AP-2 $\gamma$  is a developmentally regulated marker of testicular carcinoma in situ and germ cell tumors. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 10(24):8521–8530, 2004-12-15. ISSN 1078-0432. doi: 10.1158/1078-0432.CCR-04-1285.
- [113] S. Schäfer, J. Anschlag, D. Nettersheim, N. Haas, L. Pawig, and H. Schorle. The role of BLIMP1 and its putative downstream target TFAP2C in germ cell development and germ cell tumours. *International Journal of Andrology*, 34:e152–158; discussion e158–159, 2011-08. ISSN 1365-2605. doi: 10.1111/j.1365-2605.2011.01167.x.
- [114] J. Kang, W. Kim, S. Lee, D. Kwon, J. Chun, B. Son, E. Kim, J.-M. Lee, H. Youn, and B. Youn. TFAP2C promotes lung tumorigenesis and aggressiveness through miR-183- and miR-33a-mediated cell cycle regulation. *Oncogene*, 36(11):1585–1596, 2017-03. ISSN 1476-5594. doi: 10.1038/onc.2016.328.
- [115] Tzu-Hua Chang, Meng-Feng Tsai, Chien-Hung Gow, Shang-Gin Wu, Yi-Nan Liu, Yih-Leong Chang, Sung-Liang Yu, Hsing-Chen Tsai, Shih-Wen Lin, Yen-Wei Chen, Po-Yen Kuo, Pan-Chyr Yang, and Jin-Yuan Shih. Upregulation of microRNA-137 expression by Slug promotes tumor invasion and metastasis of non-small cell lung cancer cells through suppression of TFAP2C. *Cancer Letters*, 402:190–202, 2017-08-28. ISSN 1872-7980. doi: 10.1016/j.canlet.2017.06.002.
- [116] Hyunhee Do, Dain Kim, JiHoon Kang, Beomseok Son, Danbi Seo, HyeSook Youn, BuHyun Youn, and Wanyeon Kim. TFAP2C increases cell proliferation by downregulating GADD45B and PMAIP1 in non-small cell lung cancer cells. *Biological Research*, 52(1):35, 2019-07-11. ISSN 0717-6287. doi: 10.1186/s40659-019-0244-5.



- [117] Guangbing Xiong, Hua Huang, Mengyu Feng, Gang Yang, Suli Zheng, Lei You, Lianfang Zheng, Ya Hu, Taiping Zhang, and Yupei Zhao. MiR-10a-5p targets TFAP2C to promote gemcitabine resistance in pancreatic ductal adenocarcinoma. *Journal of Experimental & Clinical Cancer Research : CR*, 37:76, 2018-04-03. ISSN 0392-9078. doi: 10.1186/s13046-018-0739-x.
- [118] Damian Kołat, Żaneta Kałuzińska, Andrzej K. Bednarek, and Elżbieta Pluciennik. Fragile Gene WWOX Guides TFAP2A/TFAP2C-Dependent Actions Against Tumor Progression in Grade II Bladder Cancer. *Frontiers in Oncology*, 11:621060, 2021-02-25. ISSN 2234-943X. doi: 10.3389/fonc.2021.621060.
- [119] Hironobu Yamashita, Yuka I. Kawasawa, Lauren Shuman, Zongyu Zheng, Truc Tran, Vonn Walter, Joshua I. Warrick, Guoli Chen, Hikmat Al-Ahmadie, Matthew Kaag, Pak Kin Wong, Jay D. Raman, and David J. DeGraff. Repression of transcription factor AP-2 alpha by PPAR $\gamma$  reveals a novel transcriptional circuit in basal-squamous bladder cancer. *Oncogenesis*, 8(12):1–17, 2019-11-26. ISSN 2157-9024. doi: 10.1038/s41389-019-0178-3.
- [120] Xu Wang, Di Sun, Jiandong Tai, Si Chen, Miao Yu, Dong Ren, and Lei Wang. TFAP2C promotes stemness and chemotherapeutic resistance in colorectal cancer via inactivating hippo signaling pathway. *Journal of experimental & clinical cancer research: CR*, 37(1):27, 2018-02-13. ISSN 1756-9966. doi: 10.1186/s13046-018-0683-9.
- [121] Shun-Li Gao, Li-Zhong Wang, Hai-Ying Liu, Dan-Li Liu, Li-Ming Xie, and Zhi-Wei Zhang. miR-200a inhibits tumor proliferation by targeting AP-2 $\gamma$  in neuroblastoma cells. *Asian Pacific journal of cancer prevention: APJCP*, 15(11):4671–4676, 2014. ISSN 2476-762X. doi: 10.7314/apjcp.2014.15.11.4671.
- [122] Yuan Wang, Shuang Chen, Qingyuan Jiang, Jie Deng, Fuyi Cheng, Yi Lin, Lin Cheng, Yixin Ye, Xiaolei Chen, Yunqi Yao, Xiaomei Zhang, Gang Shi, Lei Dai, Xiaolan Su, Yong Peng, and Hongxin Deng. TFAP2C facilitates somatic cell reprogramming by inhibiting c-Myc-dependent apoptosis and promoting mesenchymal-to-epithelial transition. *Cell Death & Disease*, 11(6):482, 2020-06-25. ISSN 2041-4889. doi: 10.1038/s41419-020-2684-9.
- [123] Peter Kuckenber, Caroline Kubaczka, and Hubert Schorle. The role of transcription factor Tcfap2c/TFAP2C in trophectoderm development. *Reproductive Biomedicine Online*, 25(1):12–20, 2012-07. ISSN 1472-6491. doi: 10.1016/j.rbmo.2012.02.015.
- [124] Heidi J. Auman, Timothy Nottoli, Olga Lakiza, Quinton Winger, Stephanie Donaldson, and Trevor Williams. Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. *Development (Cambridge, England)*, 129(11):2733–2747, 2002-06. ISSN 0950-1991. doi: 10.1242/dev.129.11.2733.
- [125] Peter Kuckenber, Sandra Buhl, Tatiana Woynecki, prefix=van useprefix=true family=Fürden, given=Betina, Elena Tolkunova, Friederike Seiffe, Markus Moser, Alexey Tomilin, Elke Winterhager, and Hubert Schorle. The Transcription Factor TCFAP2C/AP-2 $\gamma$  Cooperates with CDX2 To Maintain Trophectoderm Formation. *Molecular and Cellular Biology*, 30(13):3310–3320, 2010-07. ISSN 0270-7306. doi: 10.1128/MCB.01215-09.
- [126] Inchul Choi, Timothy S. Carey, Catherine A. Wilson, and Jason G. Knott. Transcription factor AP-2 $\gamma$  is a core regulator of tight junction biogenesis and cavity formation during mouse early embryogenesis. *Development (Cambridge, England)*, 139(24):4623–4632, 2012-12. ISSN 1477-9129. doi: 10.1242/dev.086645.
- [127] Julie R. Hens and John J. Wysolmerski. Key stages of mammary gland development: Molecular mechanisms involved in the formation of the embryonic mammary gland. *Breast cancer research: BCR*, 7(5):220–224, 2005. ISSN 1465-542X. doi: 10.1186/bcr1306.
- [128] Nai Yang Fu, Emma Nolan, Geoffrey J. Lindeman, and Jane E. Visvader. Stem Cells and the Differentiation Hierarchy in Mammary Gland Development. *Physiological Reviews*, 100(2):489–523, 2020-04-01. ISSN 1522-1210. doi: 10.1152/physrev.00040.2018.
- [129] Hector Macias and Lindsay Hinck. Mammary Gland Development. *Wiley interdisciplinary reviews. Developmental biology*, 1(4):533–557, 2012. ISSN 1759-7684. doi: 10.1002/wdev.35.

- [130] Richard Jäger, Uwe Werling, Stephan Rimpf, Andrea Jacob, and Hubert Schorle. Transcription factor AP-2gamma stimulates proliferation and apoptosis and impairs differentiation in a transgenic model. *Molecular cancer research: MCR*, 1(12):921–929, 2003-10. ISSN 1541-7786.
- [131] Richard Jäger, Sabine Schäfer, Mathilde Hau-Liersch, and Hubert Schorle. Loss of transcription factor AP-2gamma/TFAP2C impairs branching morphogenesis of the murine mammary gland. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 239(3):1027–1033, 2010-03. ISSN 1097-0177. doi: 10.1002/dvdy.22239.
- [132] The Jackson Laboratory. MGI Statistics for the Mouse Genome Informatics database resource, 2022-03-12. URL [http://www.informatics.jax.org/mgihome/homepages/stats/all\\_stats.shtml](http://www.informatics.jax.org/mgihome/homepages/stats/all_stats.shtml).
- [133] Samuel A. Lambert, Arttu Jolma, Laura F. Campitelli, Pratyush K. Das, Yimeng Yin, Mihai Albu, Xiaoting Chen, Jussi Taipale, Timothy R. Hughes, and Matthew T. Weirauch. The Human Transcription Factors. *Cell*, 172(4):650–665, 2018-02-08. ISSN 0092-8674. doi: 10.1016/j.cell.2018.01.029.
- [134] Han Wu, Alex S. Nord, Jennifer A. Akiyama, Malak Shoukry, Veena Afzal, Edward M. Rubin, Len A. Pennacchio, and Axel Visel. Tissue-specific RNA expression marks distant-acting developmental enhancers. *PLoS genetics*, 10(9):e1004610, 2014-09. ISSN 1553-7404. doi: 10.1371/journal.pgen.1004610.
- [135] Patricia J. Wittkopp and Gizem Kalay. Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, 13(1):59–69, 2012-01. ISSN 1471-0064. doi: 10.1038/nrg3095.
- [136] Julian A. Segert, Stephen S. Gisselbrecht, and Martha L. Bulyk. Transcriptional Silencers: Driving Gene Expression with the Brakes On. *Trends in Genetics*, 37(6):514–527, 2021-06-01. ISSN 0168-9525. doi: 10.1016/j.tig.2021.02.002.
- [137] Daria Shlyueva, Gerald Stampfel, and Alexander Stark. Transcriptional enhancers: From properties to genome-wide predictions. *Nature Reviews Genetics*, 15(4):272–286, 2014-04. ISSN 1471-0064. doi: 10.1038/nrg3682.
- [138] Gráinne Barkess and Adam G West. Chromatin insulator elements: Establishing barriers to set heterochromatin boundaries. *Epigenomics*, 4(1):67–80, 2012-02. ISSN 1750-1911. doi: 10.2217/epi.11.112.
- [139] Laura A. Lettice, Simon J. H. Heaney, Lorna A. Purdie, Li Li, Philippe de Beer, Ben A. Oostra, Debbie Goode, Greg Elgar, Robert E. Hill, and Ester de Graaff. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, 12(14):1725–1735, 2003-07-15. ISSN 0964-6906. doi: 10.1093/hmg/ddg180.
- [140] Helen Ray-Jones and Mikhail Spivakov. Transcriptional enhancers and their communication with gene promoters. *Cellular and Molecular Life Sciences*, 78(19-20):6453–6485, 2021. ISSN 1420-682X. doi: 10.1007/s00018-021-03903-w.
- [141] Giuliano Crispatzu, Rizwan Rehimi, Tomas Pachano, Tore Bleckwehl, Sara Cruz-Molina, Cally Xiao, Esther Mahabir, Hisham Bazzi, and Alvaro Rada-Iglesias. The chromatin, topological and regulatory properties of pluripotency-associated poised enhancers are conserved in vivo. *Nature Communications*, 12(1):4344, 2021-07-16. ISSN 2041-1723. doi: 10.1038/s41467-021-24641-4.
- [142] Evgeny Z. Kvon, Rachel Waymack, Mario G. Elabd, and Zeba Wunderlich. Enhancer redundancy in development and disease. *Nature reviews. Genetics*, 22(5):324–336, 2021-05. ISSN 1471-0056. doi: 10.1038/s41576-020-00311-x.
- [143] Marco Osterwalder, Iros Barozzi, Virginie Tissières, Yoko Fukuda-Yuzawa, Brandon J. Mannion, Sarah Y. Afzal, Elizabeth A. Lee, Yiwen Zhu, Ingrid Plajzer-Frick, Catherine S. Pickle, Momoe Kato, Tyler H. Garvin, Quan T. Pham, Anne N. Harrington, Jennifer A. Akiyama, Veena Afzal, Javier Lopez-Rios, Diane E. Dickel, Axel Visel, and Len A. Pennacchio. Enhancer Redundancy Allows for Phenotypic Robustness in Mammalian Development. *Nature*, 554(7691):239–243, 2018-02-08. ISSN 0028-0836. doi: 10.1038/nature25461.

- [144] Puja Agrawal and Sridhar Rao. Super-Enhancers and CTCF in Early Embryonic Cell Fate Decisions. *Frontiers in Cell and Developmental Biology*, 9:653669, 2021-03-25. ISSN 2296-634X. doi: 10.3389/fcell.2021.653669.
- [145] Jill E. Moore, Michael J. Purcaro, Henry E. Pratt, Charles B. Epstein, Noam Shores, Jessika Adrian, Trupti Kawli, Carrie A. Davis, Alexander Dobin, Rajinder Kaul, Jessica Halow, Eric L. Van Nostrand, Peter Freese, David U. Gorkin, Yin Shen, Yupeng He, Mark Mackiewicz, Florencia Pauli-Behn, Brian A. Williams, Ali Mortazavi, Cheryl A. Keller, Xiao-Ou Zhang, Shaimae I. Elhajjajy, Jack Huey, Diane E. Dickel, Valentina Snetkova, Xintao Wei, Xiaofeng Wang, Juan Carlos Rivera-Mulia, Joel Rozowsky, Jing Zhang, Surya B. Chhetri, Jialing Zhang, Alec Victorsen, Kevin P. White, Axel Visel, Gene W. Yeo, Christopher B. Burge, Eric Lécuycer, David M. Gilbert, Job Dekker, John Rinn, Eric M. Mendenhall, Joseph R. Ecker, Manolis Kellis, Robert J. Klein, William S. Noble, Anshul Kundaje, Roderic Guigó, Peggy J. Farnham, J. Michael Cherry, Richard M. Myers, Bing Ren, Brenton R. Graveley, Mark B. Gerstein, Len A. Pennacchio, Michael P. Snyder, Bradley E. Bernstein, Barbara Wold, Ross C. Hardison, Thomas R. Gingeras, John A. Stamatoyannopoulos, and Zhiping Weng. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, 583(7818): 699–710, 2020-07. ISSN 1476-4687. doi: 10.1038/s41586-020-2493-4.
- [146] Ian C. Tobias, Luis E. Abatti, Sakthi D. Moorthy, Shanell Mullany, Tiegh Taylor, Nawrah Khader, Mario A. Filice, and Jennifer A. Mitchell. Transcriptional enhancers: From prediction to functional assessment on a genome-wide scale. *Genome*, 64(4):426–448, 2021-04. ISSN 0831-2796, 1480-3321. doi: 10.1139/gen-2020-0104.
- [147] Jon-Matthew Belton, Rachel Patton McCord, Johan Harmen Gibcus, Natalia Naumova, Ye Zhan, and Job Dekker. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods (San Diego, Calif.)*, 58(3):268–276, 2012-11. ISSN 1095-9130. doi: 10.1016/j.ymeth.2012.05.001.
- [148] Stefan Schoenfelder, Biola-Maria Javierre, Mayra Furlan-Magaril, Steven W. Wingett, and Peter Fraser. Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. *Journal of Visualized Experiments : JoVE*, (136):57320, 2018-06-28. ISSN 1940-087X. doi: 10.3791/57320.
- [149] Eugene V. Koonin and Kira S. Makarova. Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1772):20180087, 2019-05-13. doi: 10.1098/rstb.2018.0087.
- [150] Andrew V. Anzalone, Luke W. Koblan, and David R. Liu. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology*, 38(7):824–844, 2020-07. ISSN 1087-0156. doi: 10.1038/s41587-020-0561-9.
- [151] Nicholas S. McCarty, Alicia E. Graham, Lucie Studená, and Rodrigo Ledesma-Amaro. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nature Communications*, 11(1):1281, 2020-03-09. ISSN 2041-1723. doi: 10.1038/s41467-020-15053-x.
- [152] Matthew C. Canver, Elenoe C. Smith, Falak Sher, Luca Pinello, Neville E. Sanjana, Ophir Shalem, Diane D. Chen, Patrick G. Schupp, Divya S. Vinjamur, Sara P. Garcia, Sidinh Luc, Ryo Kurita, Yukio Nakamura, Yuko Fujiwara, Takahiro Maeda, Guo-Cheng Yuan, Feng Zhang, Stuart H. Orkin, and Daniel E. Bauer. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*, 527(7577):192–197, 2015-11. ISSN 1476-4687. doi: 10.1038/nature15521.
- [153] Neville E. Sanjana, Ophir Shalem, and Feng Zhang. Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods*, 11(8):783–784, 2014-08. ISSN 1548-7091. doi: 10.1038/nmeth.3047.
- [154] Konstantinos Tzelepis, Hiroko Koike-Yusa, Etienne De Braekeleer, Yilong Li, Emmanouil Metzakopian, Oliver M. Dovey, Annalisa Mupo, Vera Grinkevich, Meng Li, Milena Mazan, Malgorzata Gozdecka, Shuhei Ohnishi, Jonathan Cooper, Miten Patel, Thomas McKerrell, Bin Chen, Ana Filipa Domingues, Paolo Gallipoli, Sarah Teichmann, Hannes Ponstingl, Ultan McDermott, Julio Saez-Rodriguez, Brian J. P. Huntly, Francesco Iorio, Cristina Pina, George S. Vassiliou, and Kosuke Yusa. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Reports*, 17(4):1193–1205, 2016-10. ISSN 22111247. doi: 10.1016/j.celrep.2016.09.079.

- [155] Peter C. DeWeirdt, Annabel K. Sangree, Ruth E. Hanna, Kendall R. Sanson, Mudra Hegde, Christine Strand, Nicole S. Persky, and John G. Doench. Genetic screens in isogenic mammalian cell lines without single cell cloning. *Nature Communications*, 11(1):752, 2020-02-06. ISSN 2041-1723. doi: 10.1038/s41467-020-14620-6.
- [156] Julia Joung, Silvana Konermann, Jonathan S. Gootenberg, Omar O. Abudayyeh, Randall J. Platt, Mark D. Brigham, Neville E. Sanjana, and Feng Zhang. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols*, 12(4):828–863, 2017-04. ISSN 1754-2189. doi: 10.1038/nprot.2017.016.
- [157] Neville E. Sanjana, Jason Wright, Kaijie Zheng, Ophir Shalem, Pierre Fontanillas, Julia Joung, Christine Cheng, Aviv Regev, and Feng Zhang. High-resolution interrogation of functional elements in the noncoding genome. *Science*, 353(6307):1545–1549, 2016-09-30. doi: 10.1126/science.aaf7613.
- [158] Matthew C. Canver, Pratibha Tripathi, Michael J. Bullen, Moshe Olshansky, Yogesh Kumar, Lee H. Wong, Stephen J. Turner, Samuel Lessard, Luca Pinello, Stuart H. Orkin, and Partha Pratim Das. A saturating mutagenesis CRISPR-Cas9-mediated functional genomic screen identifies cis- and trans-regulatory elements of Oct4 in murine ESCs. *The Journal of Biological Chemistry*, 295(47):15797–15809, 2020-11-20. ISSN 0021-9258. doi: 10.1074/jbc.RA120.013772.
- [159] Juan J. Tena and José M. Santos-Pereira. Topologically Associating Domains and Regulatory Landscapes in Development, Evolution and Disease. *Frontiers in Cell and Developmental Biology*, 9, 2021. ISSN 2296-634X. URL <https://www.frontiersin.org/article/10.3389/fcell.2021.702787>.
- [160] Gil Ron, Yuval Globerson, Dror Moran, and Tommy Kaplan. Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nature Communications*, 8(1):2237, 2017-12. ISSN 2041-1723. doi: 10.1038/s41467-017-02386-3.
- [161] NCBI. GRCm38 - Mm10 - Genome - Assembly - NCBI. URL [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001635.20/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.20/).
- [162] Wolfram H. Gruhn and Ufuk Günesdogan. Generation of Primordial Germ Cell-like Cells on Small and Large Scales. *Methods in Molecular Biology (Clifton, N.J.)*, 2214:75–89, 2021. ISSN 1940-6029. doi: 10.1007/978-1-0716-0958-3\_6.
- [163] Thermo Fisher Scientific. Lipofectamine 2000, 2006. URL <https://www.thermofisher.com/de/de/home/references/protocols/cell-culture/transfection-protocol/lipofectamine-2000.html#procedure>.
- [164] Ufuk Günesdogan. Unpublished transfection medium protocol, 2018.
- [165] Neville Sanjana Lab. Graphical User Interface for DNA Editing Screens (GUIDES) tool, n.d.. URL <http://guides.sanjanalab.org/>.
- [166] F. Ann Ran, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11):2281–2308, 2013-11. ISSN 1754-2189. doi: 10.1038/nprot.2013.143.
- [167] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121):819–823, 2013-02. ISSN 0036-8075. doi: 10.1126/science.1231143.
- [168] Addgene. pX330-U6-Chimeric\_BB-CBh-hSpCas9, n.d.. URL <http://www.addgene.org/42230/>.
- [169] McLab. DH5-Alpha Competent E. coli Transformation User Manual, 2012-07. URL [https://www.mclab.com/mclab\\_pages/Dh5-AlphaUserManual.pdf](https://www.mclab.com/mclab_pages/Dh5-AlphaUserManual.pdf).
- [170] Macherey-Nagel. User manual Plasmid DNA purification, 2021. URL <https://www.mn-net.com/media/pdf/45/51/02/Instruction-NucleoSpin-Plasmid.pdf>.
- [171] Macherey-Nagel. User Manual NucleoSpin® Plasmid Transfection-grade, 2021. URL <https://www.mn-net.com/media/pdf/ca/39/9a/Instruction-NucleoSpin-Plasmid-Transfection-grade.pdf>.

- [172] Jamie A. Hackett, Yun Huang, Ufuk Günesdogan, Kristjan A. Gretarsson, Toshihiro Kobayashi, and M. Azim Surani. Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nature Communications*, 9(1):4292, 2018-10-16. ISSN 2041-1723. doi: 10.1038/s41467-018-06230-0.
- [173] New England BioLabs. PCR Protocol for OneTaq® DNA Polymerase (M0480) V.1. *protocols.io*, 2015-02-01. doi: dx.doi.org/10.17504/protocols.io.cq8vzv.
- [174] New England BioLabs. Ligation Protocol for NEB PCR Cloning Kit | NEB, n.d.. URL <https://international.neb.com/protocols/2013/12/27/ligation-protocol-e1202>.
- [175] Toshihiro Kobayashi, Haixin Zhang, Walfred W.C. Tang, Naoko Irie, Sarah Withey, Doris Klisch, Anastasiya Sybirna, Sabine Dietmann, David A. Contreras, Robert Webb, Cinzia Allegrucci, Ramiro Alberio, and M. Azim Surani. Principles of early human development and germ cell program from conserved model systems. *Nature*, 546(7658):416–420, 2017-06-15. ISSN 0028-0836. doi: 10.1038/nature22812.
- [176] New England BioLabs. PCR Using Q5 High-Fidelity DNA Polymerase (M0491), n.d.. URL <https://international.neb.com/protocols/2013/12/13/pcr-using-q5-high-fidelity-dna-polymerase-m0491>.
- [177] New England BioLabs. Tm Calculator, n.d.. URL <https://tmcalculator.neb.com/#!/main>.
- [178] Cold Spring Harbor. TE Buffer. *Cold Spring Harbor Protocols*, 2009(1):pdb.rec11601, 2009-01. ISSN 1940-3402. doi: 10.1101/pdb.rec11601.
- [179] Todd S. Macfarlan, Wesley D. Gifford, Shawn Driscoll, Karen Lettieri, Helen M. Rowe, Dario Bonanomi, Amy Firth, Oded Singer, Didier Trono, and Samuel L. Pfaff. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, 487(7405):57–63, 2012-07-05. ISSN 1476-4687. doi: 10.1038/nature11244.
- [180] Addgene. 2C::tdTomato Reporter, n.d.. URL <http://www.addgene.org/40281/>.
- [181] Daniel G. Gibson, Lei Young, Ray-Yuan Chuang, J. Craig Venter, Clyde A. Hutchison, and Hamilton O. Smith. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345, 2009-05. ISSN 1548-7105. doi: 10.1038/nmeth.1318.
- [182] Feng Zhang Lab. Guide Design Resources, n.d.. URL <https://zlab.bio/guide-design-resources>.
- [183] Konstantinos Anastassiadis, Jun Fu, Christoph Patsch, Shengbiao Hu, Stefanie Weidlich, Kristin Duerschke, Frank Buchholz, Frank Edenhofer, and A. Francis Stewart. Dre recombinase, like Cre, is a highly efficient site-specific recombinase in *E. coli*, mammalian cells and mice. *Disease Models & Mechanisms*, 2(9-10):508–515, 2009-09. ISSN 1754-8411. doi: 10.1242/dmm.003087.
- [184] Jan Kapuscinski. DAPI: A DNA-Specific Fluorescent Probe. *Biotechnic & Histochemistry*, 70(5):220–233, 1995-01. ISSN 1052-0295. doi: 10.3109/10520299509108199.
- [185] Martin Hooper, Kate Hardy, Alan Handyside, Susan Hunter, and Marilyn Monk. HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature*, 326(6110):292–295, 1987-03. ISSN 0028-0836. doi: 10.1038/326292a0.
- [186] Qiagen. QuantiTect® Reverse Transcription Handbook, 2009.
- [187] Smith College. Analyzing your QRT-PCR Data, 2015.
- [188] Petra Hajkova. Unpublished staining protocol, n.d.
- [189] Ismael C. Gomes, Mariana Acquarone, Renata de Moraes Maciel, Rafael Bierig Erlich, and Stevens K. Rehen. Analysis of Pluripotent Stem Cells by using Cryosections of Embryoid Bodies. *Journal of Visualized Experiments*, (46), 2010-12. ISSN 1940-087X. doi: 10.3791/2344.
- [190] Lena Wartosch. Unpublished protein lysate generation protocol, n.d.

- [191] Invitrogen. Qubit™ Protein Assay Kits, 2022-02-08. URL [https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FQubit\\_Protein\\_Assay\\_UG.pdf](https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FQubit_Protein_Assay_UG.pdf).
- [192] Cold Spring Harbor. Transfer buffer for western blotting. *Cold Spring Harbor Protocols*, 2006(4):pdb.rec10629, 2006-09. ISSN 1940-3402. doi: 10.1101/pdb.rec10629.
- [193] Bio-Rad Laboratories. Mini Trans-Blot Electrophoretic Transfer Cell Instruction Manual, n.d. URL <https://www.bio-rad.com/webroot/web/pdf/lsr/literature/M1703930.pdf>.
- [194] Agilent. Comparison of RIN and RQN for the Agilent 2100 Bioanalyzer and the Fragment Analyzer Systems, 2020-04-01.
- [195] Life Technologies. Qubit™ RNA HS Assay Kits, 2015-02-16. URL [https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FQubit\\_RNA\\_HS\\_Assay\\_UG.pdf](https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FQubit_RNA_HS_Assay_UG.pdf).
- [196] New England BioLabs. Protocol for use with NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490), 2017. URL <https://international.neb.com/protocols/2017/02/05/protocol-for-use-with-nebnext-poly-a-mrna-magnetic-isolation-module-e7490-and-nebnext-ultra-ii-di>.
- [197] Beckman Coulter. Instructions for Use RNAClean XP, 2020-05.
- [198] Enis Afgan, Dannon Baker, B er enice Batut, prefix=van den useprefix=false family=Beek, given=Marius, Dave Bouvier, Martin ech, John Chilton, Dave Clements, Nate Coraor, Bj orn A. Gr uning, Aysam Guerler, Jennifer Hillman-Jackson, Saskia Hiltmann, Vahid Jalili, Helena Rasche, Nicola Soranzo, Jeremy Goecks, James Taylor, Anton Nekrutenko, and Daniel Blankenberg. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(W1):W537–W544, 2018-07. ISSN 0305-1048. doi: 10.1093/nar/gky379.
- [199] Simon Andrews. FastQC, 2015-10-09. URL <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- [200] Philip Ewels, M ans Magnusson, Sverker Lundin, and Max K aller. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048, 2016-10. ISSN 1367-4803. doi: 10.1093/bioinformatics/btw354.
- [201] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10, 2011-05. ISSN 2226-6089. doi: 10.14806/ej.17.1.200.
- [202] Daehwan Kim, Ben Langmead, and Steven L. Salzberg. HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4):357–360, 2015-04. ISSN 1548-7091. doi: 10.1038/nmeth.3317.
- [203] Broad Institute. Picard Toolkit, 2019. URL <http://broadinstitute.github.io/picard/>.
- [204] Y. Liao, G. K. Smyth, and W. Shi. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930, 2014-04. ISSN 1367-4803. doi: 10.1093/bioinformatics/btt656.
- [205] Mark Dunning. annotateMyIDs, 2017-03-23. URL <https://github.com/markdunning/galaxy-annotateMyIDs>.
- [206] Michael I. Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, 2014-12. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8.
- [207] Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz, Bill Venables, and Tal Galili. Gplots, n.d. URL <https://cran.r-project.org/web/packages/gplots/index.html>.

- [208] Matthew D. Young, Matthew J. Wakefield, Gordon K. Smyth, and Alicia Oshlack. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology*, 11(2):R14, 2010. ISSN 1465-6906. doi: 10.1186/gb-2010-11-2-r14.
- [209] Federico Marini and Harald Binder. pcaExplorer: An R/Bioconductor package for interacting with RNA-seq principal components. *BMC Bioinformatics*, 20(1):331, 2019-12. ISSN 1471-2105. doi: 10.1186/s12859-019-2879-1.
- [210] Ligu Wang, Shengqin Wang, and Wei Li. RSeQC: Quality control of RNA-seq experiments. *Bioinformatics*, 28(16):2184–2185, 2012-08. ISSN 1460-2059. doi: 10.1093/bioinformatics/bts356.
- [211] Petr Danecek, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, Thomas Keane, Shane A. McCarthy, Robert M. Davies, and Heng Li. Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), 2021-01. ISSN 2047-217X. doi: 10.1093/gigascience/giab008.
- [212] Helga Thorvaldsdóttir, James T. Robinson, and Jill P. Mesirov. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2):178–192, 2013-03-01. ISSN 1467-5463. doi: 10.1093/bib/bbs017.
- [213] Fidel Ramírez, Devon P. Ryan, Björn Grüning, Vivek Bhardwaj, Fabian Kilpert, Andreas S. Richter, Steffen Heyne, Friederike Dündar, and Thomas Manke. deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1):W160–W165, 2016-07. ISSN 0305-1048. doi: 10.1093/nar/gkw257.
- [214] Jason D. Buenrostro, Beijing Wu, Howard Y. Chang, and William J. Greenleaf. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology*, 109(1), 2015-01. ISSN 1934-3639. doi: 10.1002/0471142727.mb2129s109.
- [215] Felix Krueger. TrimGalore, n.d.
- [216] Ben Langmead and Steven L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359, 2012-04. ISSN 1548-7091. doi: 10.1038/nmeth.1923.
- [217] Tao Liu. MACS2, 2019-09-06. URL <https://github.com/macs3-project/MACS/releases?page=2>.
- [218] R Core Team. R, 2021. URL <https://www.R-project.org/>.
- [219] Ruijie Liu, Aliaksei Z. Holik, Shian Su, Natasha Jansz, Kelan Chen, Huei San Leong, Marnie E. Blewitt, Marie-Liesse Asselin-Labat, Gordon K. Smyth, and Matthew E. Ritchie. Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Research*, 43(15):e97, 2015-09-03. ISSN 1362-4962. doi: 10.1093/nar/gkv412.
- [220] Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1):139–140, 2010-01-01. ISSN 1367-4811. doi: 10.1093/bioinformatics/btp616.
- [221] Sven Heinz, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, 38(4):576–589, 2010-05-28. ISSN 1097-4164. doi: 10.1016/j.molcel.2010.05.004.
- [222] Jake R Conway, Alexander Lex, and Nils Gehlenborg. UpSetR: An R Package for the Visualization of Intersecting Sets and Their Properties. *Bioinformatics*, 33(18):2938–2940, 2017-09-15. ISSN 1367-4803. doi: 10.1093/bioinformatics/btx364.
- [223] Stephen Kelly. HOMER\_mergePeaks\_venn\_UpSetR, 2022-04-25T13:54:10Z. URL [https://github.com/stevekm/Bioinformatics/blob/776c420efac851c6780ce573939fb6610a3b9ae8/HOMER\\_mergePeaks\\_pipeline/HOMER\\_mergePeaks\\_venn\\_UpSetR/multi\\_peaks\\_UpSet\\_plot.R](https://github.com/stevekm/Bioinformatics/blob/776c420efac851c6780ce573939fb6610a3b9ae8/HOMER_mergePeaks_pipeline/HOMER_mergePeaks_venn_UpSetR/multi_peaks_UpSet_plot.R).

- [224] Feng Zhang Lab. Count\_spacers.py, 2017-03-23. URL [https://github.com/fengzhanglab/Screening\\_Protocols\\_manuscript/blob/master/count\\_spacers.py](https://github.com/fengzhanglab/Screening_Protocols_manuscript/blob/master/count_spacers.py).
- [225] Benchling. CRISPR Guide RNA Design, n.d. URL <https://www.thermofisher.com/de/de/home/references/protocols/cell-culture/transfection-protocol/lipofectamine-2000.html#procedure>.
- [226] Ophir Shalem, Neville E. Sanjana, Ella Hartenian, Xi Shi, David A. Scott, Tarjei S. Mikkelsen, Dirk Heckl, Benjamin L. Ebert, David E. Root, John G. Doench, and Feng Zhang. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science*, 343(6166):84–87, 2014-01. ISSN 0036-8075. doi: 10.1126/science.1247005.
- [227] Addgene. pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W, n.d.. URL <https://www.addgene.org/67974/>.
- [228] QIAGEN. QIAquick Spin Handbook, 2020-01. URL <https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/dna-clean-up/qiaquick-nucleotide-removal-kit/>.
- [229] Lonza Group AG. Section VII: Separation of DNA in Polyacrylamide Gels, n.d. URL <https://lonza.picturepark.com/Website/?Action=downloadAsset&AssetId=30899>.
- [230] Macherey-Nagel. User manual PCR Clean-Up and Gel Extraction, 2021. URL <https://www.mn-net.com/media/pdf/02/1a/74/Instruction-NucleoSpin-Gel-and-PCR-Clean-up.pdf>.
- [231] New England BioLabs. Ligation Protocol WITH T4 DNA Ligase (M0202) V.1, 2015. URL [https://www.protocols.io/view/Ligation-Protocol-WITH-T4-DNA-Ligase-M0202-imss4v?version\\_warning=no](https://www.protocols.io/view/Ligation-Protocol-WITH-T4-DNA-Ligase-M0202-imss4v?version_warning=no).
- [232] New England BioLabs. Electroporation Protocol (C2986) V.1, 2015. URL <https://www.protocols.io/view/Electroporation-Protocol-C2986-imsv3v>.
- [233] Bio-Rad. Gene Pulser Xcell™ Electroporation System Instruction Manual, n.d. URL <https://www.bio-rad.com/webroot/web/pdf/lsr/literature/4006217A.pdf>.
- [234] QIAGEN. Quick-Start Protocol EndoFree® Plasmid Maxi Kit, 2016-03.
- [235] QIAGEN. MinElute® Handbook, 2020-01.
- [236] Wei Li, Han Xu, Tengfei Xiao, Le Cong, Michael I. Love, Feng Zhang, Rafael A. Irizarry, Jun S. Liu, Myles Brown, and X. Shirley Liu. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology*, 15(12):554, 2014-12-05. ISSN 1474-760X. doi: 10.1186/s13059-014-0554-4.
- [237] Addgene. pMD2.G, n.d.. URL <https://www.addgene.org/12259/>.
- [238] Addgene. psPAX2, n.d.. URL <https://www.addgene.org/12260/>.
- [239] Robert William Figliozzi, Feng Chen, Albert Chi, and Shao-Chung Victor Hsia. Using the inverse Poisson distribution to calculate multiplicity of infection and viral replication by a high-throughput fluorescent imaging system. *Virologica Sinica*, 31(2):180, 2016-04. doi: 10.1007/s12250-015-3662-8.
- [240] Zymo Research. Quick-DNA™ Miniprep Plus Kit, n.d.
- [241] Wei Li, Johannes Köster, Han Xu, Chen-Hao Chen, Tengfei Xiao, Jun S. Liu, Myles Brown, and X. Shirley Liu. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biology*, 16(1):281, 2015-12-16. ISSN 1474-760X. doi: 10.1186/s13059-015-0843-6.
- [242] Integrated DNA Technologies. CRISPR-Cas9 Guide RNA Design Checker | IDT, n.d. URL [https://eu.idtdna.com/site/order/designtool/index/CRISPR\\_SEQUENCE](https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE).
- [243] W. James Kent, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and David Haussler. The human genome browser at UCSC. *Genome Research*, 12(6):996–1006, 2002-06. ISSN 1088-9051. doi: 10.1101/gr.229102.



- [244] W. James Kent. BLAT—The BLAST-Like Alignment Tool. *Genome Research*, 12(4):656–664, 2002-04-01. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.229202.
- [245] Tim Hulsen. DeepVenn, 2022. URL <https://www.deepvenn.com>.
- [246] Tim Hulsen, prefix=de useprefix=true family=Vlieg, given=Jacob, and Wynand Alkema. BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, 9(1):488, 2008-10-16. ISSN 1471-2164. doi: 10.1186/1471-2164-9-488.
- [247] Koen Van den Berge, Katharina M. Hembach, Charlotte Soneson, Simone Tiberi, Lieven Clement, Michael I. Love, Rob Patro, and Mark D. Robinson. RNA Sequencing Data: Hitchhiker’s Guide to Expression Analysis. *Annual Review of Biomedical Data Science*, 2(1):139–173, 2019. doi: 10.1146/annurev-biodatasci-072018-021255.
- [248] Courtney W. Hanna and Gavin Kelsey. Genomic Imprinting beyond DNA Methylation: A Role for Maternal Histones. *Genome Biology*, 18(1):177, 2017-09-19. ISSN 1474-760X. doi: 10.1186/s13059-017-1317-9.
- [249] Renumathy Dhanasekaran, Anja Deutzmann, Wadie D. Mahauad-Fernandez, Aida S. Hansen, Arvin M. Gouw, and Dean W. Felsher. The MYC oncogene — the grand orchestrator of cancer growth and immune evasion. *Nature Reviews Clinical Oncology*, 19(1):23–36, 2022-01. ISSN 1759-4782. doi: 10.1038/s41571-021-00549-2.
- [250] Liang Wang, Xiangsheng Zuo, Keping Xie, and Daoyan Wei. The Role of CD44 and Cancer Stem Cells. *Methods in Molecular Biology (Clifton, N.J.)*, 1692:31–42, 2018. ISSN 1940-6029. doi: 10.1007/978-1-4939-7401-6\_3.
- [251] Mohammed I. Ahmed, Salah Elias, Arne W. Mould, Elizabeth K. Bikoff, and Elizabeth J. Robertson. The transcriptional repressor Blimp1 is expressed in rare luminal progenitors and is essential for mammary gland development. *Development (Cambridge, England)*, 143(10):1663–1673, 2016-05-15. ISSN 1477-9129. doi: 10.1242/dev.136358.
- [252] Aline Wuidart, Alejandro Sifrim, Marco Fioramonti, Shigeru Matsumura, Audrey Brisebarre, Daniel Brown, Alessia Centonze, Anne Dannau, Christine Dubois, Alexandra Van Keymeulen, Thierry Voet, and Cédric Blanpain. Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nature cell biology*, 20(6):666–676, 2018-06. ISSN 1465-7392. doi: 10.1038/s41556-018-0095-2.
- [253] Aurelio Balsalobre and Jacques Drouin. Pioneer Factors as Master Regulators of the Epigenome and Cell Fate. *Nature Reviews Molecular Cell Biology*, 2022-03-09. ISSN 1471-0072, 1471-0080. doi: 10.1038/s41580-022-00464-z.
- [254] Lilien N. Voong, Liqun Xi, Ji-Ping Wang, and Xiaozhong Wang. Genome-Wide Mapping of the Nucleosome Landscape by Micrococcal Nuclease and Chemical Mapping. *Trends in genetics : TIG*, 33(8):495–507, 2017-08. ISSN 0168-9525. doi: 10.1016/j.tig.2017.05.007.
- [255] Feng Yan, David R. Powell, David J. Curtis, and Nicholas C. Wong. From reads to insight: A hitchhiker’s guide to ATAC-seq data analysis. *Genome Biology*, 21(1):22, 2020-02-03. ISSN 1474-760X. doi: 10.1186/s13059-020-1929-3.
- [256] Robert C. McLeay and Timothy L. Bailey. Motif Enrichment Analysis: A unified framework and an evaluation on CHIP data. *BMC Bioinformatics*, 11(1):165, 2010. ISSN 1471-2105. doi: 10.1186/1471-2105-11-165.
- [257] William A. Pastor, Wanlu Liu, Di Chen, Jamie Ho, Rachel Kim, Timothy J. Hunt, Anastasia Lukianchikov, Xiaodong Liu, Jose M. Polo, Steven E. Jacobsen, and Amander T. Clark. TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nature Cell Biology*, 20(5):553–564, 2018-05. ISSN 1476-4679. doi: 10.1038/s41556-018-0089-0.
- [258] Samuel Rivero-Hinojosa, Elena M. Pugacheva, Sungyun Kang, Claudia Fabiola Méndez-Catalá, Alexander L. Kovalchuk, Alexander V. Strunnikov, Dmitri Loukinov, Jeannie T. Lee, and Victor V. Lobanenkov. The combined action of CTCF and its testis-specific paralog BORIS is essential for spermatogenesis. *Nature Communications*, 12(1):3846, 2021-06-22. ISSN 2041-1723. doi: 10.1038/s41467-021-24140-6.

- [259] Roy Drissen, Robert-Jan Palstra, Nynke Gillemans, Erik Splinter, Frank Grosveld, Sjaak Philipsen, and prefix=de useprefix=true family=Laat, given=Wouter. The active spatial organization of the  $\beta$ -globin locus requires the transcription factor EKLf. *Genes & Development*, 18(20):2485–2490, 2004-10-15. ISSN 0890-9369. doi: 10.1101/gad.317004.
- [260] Andrew J. Oldfield, Telmo Henriques, Dharendra Kumar, Adam B. Burkholder, Senthilkumar Cinghu, Damien Paulet, Brian D. Bennett, Pengyi Yang, Benjamin S. Scruggs, Christopher A. Lavender, Eric Rivals, Karen Adelman, and Raja Jothi. NF-Y controls fidelity of transcription initiation at gene promoters through maintenance of the nucleosome-depleted region. *Nature Communications*, 10(1):3072, 2019-07-11. ISSN 2041-1723. doi: 10.1038/s41467-019-10905-7.
- [261] Tanveer Sharif, Cathleen Dai, Emma Martell, Mohammad Saleh Ghassemi-Rad, Mark Robert Hanes, Patrick J. Murphy, Barry E. Kennedy, Chitra Venugopal, Minomi Subapanditha, Carman A. Giacomantonio, Paola Marcato, Sheila K. Singh, and Shashi Gujar. TAp73 Modifies Metabolism and Positively Regulates Growth of Cancer Stem-Like Cells in a Redox-Sensitive Manner. *Clinical Cancer Research*, 25(6):2001–2017, 2019-03-15. ISSN 1078-0432. doi: 10.1158/1078-0432.CCR-17-3177.
- [262] Yunbo Qiao, Yue Zhu, Nengyin Sheng, Jun Chen, Ran Tao, Qingqing Zhu, Ting Zhang, Cheng Qian, and Naihe Jing. AP2 $\gamma$  regulates neural and epidermal development downstream of the BMP pathway at early stages of ectodermal patterning. *Cell Research*, 22(11):1546–1561, 2012-11. ISSN 1748-7838. doi: 10.1038/cr.2012.122.
- [263] Ziqing Liu, Olivia Chen, J. Blake Joseph Wall, Michael Zheng, Yang Zhou, Li Wang, Haley Ruth Vaseghi, Li Qian, and Jiandong Liu. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific Reports*, 7(1):2193, 2017-12. ISSN 2045-2322. doi: 10.1038/s41598-017-02460-2.
- [264] B. Salomon, S. Maury, L. Loubière, M. Caruso, R. Onclercq, and D. Klatzmann. A truncated herpes simplex virus thymidine kinase phosphorylates thymidine and nucleoside analogs and does not cause sterility in transgenic mice. *Molecular and Cellular Biology*, 15(10):5322–5328, 1995-10. ISSN 0270-7306. doi: 10.1128/MCB.15.10.5322.
- [265] Jesse R. Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, 2012-04-11. ISSN 1476-4687. doi: 10.1038/nature11082.
- [266] Stefan Schönfelder. Unpublished work, 2017.
- [267] Nicola A. Kearns, Hannah Pham, Barbara Tabak, Ryan M. Genga, Noah J. Silverstein, Manuel Garber, and René Maehr. Functional annotation of native enhancers with a Cas9–histone demethylase fusion. *Nature Methods*, 12(5):401–403, 2015-05. ISSN 1548-7091. doi: 10.1038/nmeth.3325.
- [268] Tobias Otto and Piotr Sicinski. Cell cycle proteins as promising targets in cancer therapy. *Nature reviews. Cancer*, 17(2):93–115, 2017-01-27. ISSN 1474-175X. doi: 10.1038/nrc.2016.138.
- [269] Wen Yang, Peter L Hsu, Fan Yang, Jae-Eun Song, and Gabriele Varani. Reconstitution of the CstF complex unveils a regulatory role for CstF-50 in recognition of 3'-end processing signals. *Nucleic Acids Research*, 46(2):493–503, 2018-01-25. ISSN 0305-1048. doi: 10.1093/nar/gkx1177.
- [270] J. Ramesh Babu, Karthik B. Jeganathan, Darren J. Baker, Xiaosheng Wu, Ningling Kang-Decker, and prefix=van useprefix=true family=Deursen, given=Jan M. Rael is an essential mitotic checkpoint regulator that cooperates with Bub3 to prevent chromosome missegregation. *The Journal of Cell Biology*, 160(3):341–353, 2003-02-03. ISSN 0021-9525. doi: 10.1083/jcb.200211048.
- [271] Yuta Ando, Kennedy Omondi Okeyo, Junko Sunaga, and Taiji Adachi. Edge-localized alteration in pluripotency state of mouse ES cells forming topography-confined layers on designed mesh substrates. *Stem Cell Research*, 53:102352, 2021-05-01. ISSN 1873-5061. doi: 10.1016/j.scr.2021.102352.

- [272] Dennis Van Hoof, Stefan R. Braam, Wilma Dormeyer, Dorien Ward-van Oostwaard, Albert J.R. Heck, Jeroen Krijgsveld, and Christine L. Mummery. Feeder-Free Monolayer Cultures of Human Embryonic Stem Cells Express an Epithelial Plasma Membrane Protein Profile. *STEM CELLS*, 26(11):2777–2781, 2008. ISSN 1549-4918. doi: 10.1634/stemcells.2008-0365.
- [273] Danielle Park, Esther Wershof, Stefan Boeing, Anna Labernadie, Robert P. Jenkins, Samantha George, Xavier Trepas, Paul A. Bates, and Erik Sahai. Extracellular matrix anisotropy is determined by TFAP2C-dependent regulation of cell collisions. *Nature Materials*, 19(2):227–238, 2020-02. ISSN 1476-1122. doi: 10.1038/s41563-019-0504-3.
- [274] Raghu Kalluri and Robert A. Weinberg. The basics of epithelial-mesenchymal transition. *The Journal of Clinical Investigation*, 119(6):1420–1428, 2009-06-01. ISSN 0021-9738. doi: 10.1172/JCI39104.
- [275] Christopher B. Mulholland, Atsuya Nishiyama, Joel Ryan, Ryohei Nakamura, Merve Yiğit, Ivo M. Glück, Carina Trummer, Weihua Qin, Michael D. Bartoschek, Franziska R. Traube, Edris Parsa, Enes Ugur, Miha Modic, Aishwarya Acharya, Paul Stolz, Christoph Ziegenhain, Michael Wierer, Wolfgang Enard, Thomas Carell, Don C. Lamb, Hiroyuki Takeda, Makoto Nakanishi, Sebastian Bultmann, and Heinrich Leonhardt. Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive DNA demethylation in mammals. *Nature Communications*, 11(1):5972, 2020-11-24. ISSN 2041-1723. doi: 10.1038/s41467-020-19603-1.
- [276] Toshinobu Nakamura, Yoshikazu Arai, Hiroki Umehara, Masaaki Masuhara, Tohru Kimura, Hisaaki Taniguchi, Toshihiro Sekimoto, Masahito Ikawa, Yoshihiro Yoneda, Masaru Okabe, Satoshi Tanaka, Kunio Shiota, and Toru Nakano. PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature Cell Biology*, 9(1):64–71, 2007-01. ISSN 1476-4679. doi: 10.1038/ncb1519.
- [277] Rasoul Godini, Keyvan Karami, and Hossein Fallahi. Genome imprinting in stem cells: A mini-review. *Gene Expression Patterns*, 34:119063, 2019-12-01. ISSN 1567-133X. doi: 10.1016/j.gep.2019.119063.
- [278] Xingbo Xu, Lukasz Smorag, Toshinobu Nakamura, Tohru Kimura, Ralf Dressel, Antje Fitzner, Xiaoying Tan, Matthias Linke, Ulrich Zechner, Wolfgang Engel, and D. V. Krishna Pantakani. Dppa3 expression is critical for generation of fully reprogrammed iPSCs and maintenance of Dlk1-Dio3 imprinting. *Nature Communications*, 6(1):6008, 2015-01-23. ISSN 2041-1723. doi: 10.1038/ncomms7008.
- [279] Maria Schacker, Yi-Han Cheng, Melanie Eckersley-Maslin, Richard Michael Snaith, and William Henry Colledge. Hypermethylation and reduced expression of Gtl2, Rian and Mirg at the Dlk1-Dio3 imprinted locus as a marker for poor developmental potential of mouse embryonic stem cells. *Stem Cell Research*, 48:101931, 2020-10-01. ISSN 1873-5061. doi: 10.1016/j.scr.2020.101931.
- [280] Milena Stevanovic, Danijela Drakulic, Andrijana Lazic, Danijela Stanisavljevic Ninkovic, Marija Schwirtlich, and Marija Mojsin. SOX Transcription Factors as Important Regulators of Neuronal and Glial Differentiation During Nervous System Development and Adult Neurogenesis. *Frontiers in Molecular Neuroscience*, 14, 2021. ISSN 1662-5099. URL <https://www.frontiersin.org/article/10.3389/fnmol.2021.654031>.
- [281] Yuhang Zhou, Tingting Huang, Alfred S. L. Cheng, Jun Yu, Wei Kang, and Ka Fai To. The TEAD Family and Its Oncogenic Role in Promoting Tumorigenesis. *International Journal of Molecular Sciences*, 17(1):138, 2016-01-21. ISSN 1422-0067. doi: 10.3390/ijms17010138.
- [282] Xiaoxiao Wang, Yunlong Xiang, Yang Yu, Ran Wang, Yu Zhang, Qianhua Xu, Hao Sun, Zhen-Ao Zhao, Xiangxiang Jiang, Xiaoqing Wang, Xukun Lu, Dandan Qin, Yujun Quan, Jiaqi Zhang, Ng Shyh-Chang, Hongmei Wang, Naihe Jing, Wei Xie, and Lei Li. Formative Pluripotent Stem Cells Show Features of Epiblast Cells Poised for Gastrulation. *Cell Research*, 31(5):526–541, 2021-05. ISSN 1748-7838. doi: 10.1038/s41422-021-00477-x.
- [283] Vikas Malik, Dennis Zimmer, and Ralf Jauch. Diversity among POU transcription factors in chromatin recognition and cell fate reprogramming. *Cellular and Molecular Life Sciences*, 75(9):1587–1612, 2018-05-01. ISSN 1420-9071. doi: 10.1007/s00018-018-2748-5.

- [284] Guilai Shi and Ying Jin. Role of Oct4 in Maintaining and Regaining Stem Cell Pluripotency. *Stem Cell Research & Therapy*, 1(5):39, 2010-12-14. ISSN 1757-6512. doi: 10.1186/scrt39.
- [285] Christa Buecker, Rajini Srinivasan, Zhixiang Wu, Eliezer Calo, Dario Acampora, Tiago Faial, Antonio Simeone, Minjia Tan, Tomasz Swigut, and Joanna Wysocka. Reorganization of Enhancer Patterns in Transition from Naïve to Primed Pluripotency. *Cell stem cell*, 14(6):838–853, 2014-06-05. ISSN 1934-5909. doi: 10.1016/j.stem.2014.04.003.
- [286] Sascha Röth, Luke J. Fulcher, and Gopal P. Sapkota. Advances in targeted degradation of endogenous proteins. *Cellular and Molecular Life Sciences*, 76(14):2761–2777, 2019. ISSN 1420-682X. doi: 10.1007/s00018-019-03112-6.
- [287] Julie R. Hens, Pamela Dann, Jian-Ping Zhang, Stephen Harris, Gertraud W. Robinson, and John Wysolmerski. BMP4 and PTHrP interact to stimulate ductal outgrowth during embryonic mammary development and to inhibit hair follicle induction. *Development*, 134(6):1221–1230, 2007-03-15. ISSN 0950-1991. doi: 10.1242/dev.000182.
- [288] A. Hollnagel, V. Oehlmann, J. Heymer, U. Rütther, and A. Nordheim. Id genes are direct targets of bone morphogenetic protein induction in embryonic stem cells. *The Journal of Biological Chemistry*, 274(28):19838–19845, 1999-07-09. ISSN 0021-9258. doi: 10.1074/jbc.274.28.19838.
- [289] Salah Elias, Marc A. Morgan, Elizabeth K. Bikoff, and Elizabeth J. Robertson. Long-lived unipotent Blimp1-positive luminal stem cells drive mammary gland organogenesis throughout adult life. *Nature Communications*, 8(1):1714, 2017-11-20. ISSN 2041-1723. doi: 10.1038/s41467-017-01971-w.
- [290] Yoana Dimitrova, Andreas J. Gruber, Nitish Mittal, Souvik Ghosh, Beatrice Dimitriadis, Daniel Mathow, William Aaron Grandy, Gerhard Christofori, and Mihaela Zavolan. TFAP2A is a component of the ZEB1/2 network that regulates TGF $\beta$ 1-induced epithelial to mesenchymal transition. *Biology Direct*, 12(1):8, 2017-04-17. ISSN 1745-6150. doi: 10.1186/s13062-017-0180-7.
- [291] Simone Mesman, prefix=van useprefix=true family=Hooft, given=Johannes A., and Marten P. Smidt. Mest/Peg1 Is Essential for the Development and Maintenance of a SNc Neuronal Subset. *Frontiers in Molecular Neuroscience*, 9, 2017. ISSN 1662-5099. URL <https://www.frontiersin.org/article/10.3389/fnmol.2016.00166>.
- [292] Maryam Keshavarz and Diethard Tautz. The imprinted lncRNA Peg13 regulates sexual preference and the sex-specific brain transcriptome in mice. *Proceedings of the National Academy of Sciences*, 118(10):e2022172118, 2021-03-09. doi: 10.1073/pnas.2022172118.
- [293] Luisa Pinto, Daniela Drechsel, Marie-Theres Schmid, Jovica Ninkovic, Martin Irmeler, Monika S. Brill, Laura Restani, Laura Gianfranceschi, Chiara Cerri, Susanne N. Weber, Victor Tarabykin, Kristin Baer, François Guillemot, Johannes Beckers, Nada Zecevic, Colette Dehay, Matteo Caleo, Hubert Schorle, and Magdalena Götz. AP2 $\gamma$  regulates basal progenitor fate in a region- and layer-specific manner in the developing cortex. *Nature Neuroscience*, 12(10):1229–1237, 2009-10. ISSN 1546-1726. doi: 10.1038/nn.2399.
- [294] Inma Ponte, Marta Andrés, Albert Jordan, and Alicia Roque. Towards understanding the Regulation of Histone H1 Somatic Subtypes with OMICs. *Journal of Molecular Biology*, 433(2):166734, 2021-01-22. ISSN 1089-8638. doi: 10.1016/j.jmb.2020.166734.
- [295] Carlo Maria Di Liegro, Gabriella Schiera, and Italia Di Liegro. H1.0 Linker Histone as an Epigenetic Regulator of Cell Proliferation and Differentiation. *Genes*, 9(6):310, 2018-06-20. ISSN 2073-4425. doi: 10.3390/genes9060310.
- [296] Yang Yang, Liwei Zhang, Chaoyang Xiong, Jun Chen, Li Wang, Zengqi Wen, Juan Yu, Ping Chen, Yanhui Xu, Jingji Jin, Yong Cai, and Guohong Li. HIRA complex presets transcriptional potential through coordinating depositions of the histone variants H3.3 and H2A.Z on the poised genes in mESCs. *Nucleic Acids Research*, 50(1):191–206, 2021-12-10. ISSN 0305-1048. doi: 10.1093/nar/gkab1221.

- [297] Hsu-Hsin Chen, Maaïke Welling, Donald B. Bloch, Javier Muñoz, Edwin Mientjes, Xinjie Chen, Cody Tramp, Jie Wu, Akiko Yabuuchi, Yu-Fen Chou, Christa Buecker, Adrian Krainer, Rob Willemsen, Albert J. Heck, and Niels Geijsen. DAZL Limits Pluripotency, Differentiation, and Apoptosis in Developing Primordial Germ Cells. *Stem Cell Reports*, 3(5):892–904, 2014-10-11. ISSN 2213-6711. doi: 10.1016/j.stemcr.2014.09.003.
- [298] Maaïke Welling, Hsu-Hsin Chen, Javier Muñoz, Michael U. Musheev, Lennart Kester, Jan Philipp Junker, Nikolai Mischerikow, Mandana Arbab, Ewart Kuijk, Lev Silberstein, Peter V. Kharchenko, Mieke Geens, Christof Niehrs, prefix=van de useprefix=false family=Velde, given=Hilde, prefix=van useprefix=false family=Oudenaarden, given=Alexander, Albert JR Heck, and Niels Geijsen. DAZL regulates Tet1 translation in murine embryonic stem cells. *EMBO reports*, 16(7):791–802, 2015-07-01. ISSN 1469-3178. doi: 10.15252/embr.201540538.
- [299] Sigrid Eckardt, K John McLaughlin, and Holger Willenbring. Mouse chimeras as a system to investigate development, cell and tissue function, disease mechanisms and organ regeneration. *Cell Cycle*, 10(13):2091–2099, 2011-07-01. ISSN 1538-4101. doi: 10.4161/cc.10.13.16360.
- [300] Christiane Buta, Robert David, Ralf Dressel, Mia Emgård, Christiane Fuchs, Ulrike Gross, Lyn Healy, Jürgen Hescheler, Roman Kolar, Ulrich Martin, Harald Mikkers, Franz-Josef Müller, Rebekka K. Schneider, Andrea E. M. Seiler, Horst Spielmann, and Georg Weitzer. Reconsidering pluripotency tests: Do we still need teratoma assays? *Stem Cell Research*, 11(1):552–562, 2013-07-01. ISSN 1873-5061. doi: 10.1016/j.scr.2013.03.001.
- [301] Leonardo Beccari, Naomi Moris, Mehmet Girgin, David A. Turner, Peter Baillie-Johnson, Anne-Catherine Cossy, Matthias P. Lutolf, Denis Duboule, and Alfonso Martinez Arias. Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature*, 562(7726):272–276, 2018-10. ISSN 1476-4687. doi: 10.1038/s41586-018-0578-0.
- [302] L. A. Smith, A. Hidalgo Aguilar, D. D. G. Owens, R. H. Quelch, E. Knight, and S. A. Przyborski. Using Advanced Cell Culture Techniques to Differentiate Pluripotent Stem Cells and Recreate Tissue Structures Representative of Teratoma Xenografts. *Frontiers in Cell and Developmental Biology*, 9, 2021. ISSN 2296-634X. URL <https://www.frontiersin.org/article/10.3389/fcell.2021.667246>.
- [303] Siqin Bao, Fuchou Tang, Xihe Li, Katsuhiko Hayashi, Astrid Gillich, Kaiqin Lao, and M. Azim Surani. Epigenetic reversion of postimplantation epiblast cells to pluripotent embryonic stem cells. *Nature*, 461(7268):10.1038/nature08534, 2009-10-29. ISSN 0028-0836. doi: 10.1038/nature08534.
- [304] Bing Zhang, Pat Metharom, Howard Jullie, Kay AO Ellem, Geoff Cleghorn, Malcolm J West, and Ming Q Wei. The significance of controlled conditions in lentiviral vector titration and in the use of multiplicity of infection (MOI) for predicting gene transfer events. *Genetic Vaccines and Therapy*, 2:6, 2004-08-04. ISSN 1479-0556. doi: 10.1186/1479-0556-2-6.
- [305] Yarui Diao, Bin Li, Zhipeng Meng, Inkyung Jung, Ah Young Lee, Jesse Dixon, Lenka Maliskova, Kun-Liang Guan, Yin Shen, and Bing Ren. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Research*, 26(3):397–405, 2016-03. ISSN 1549-5469. doi: 10.1101/gr.197152.115.
- [306] Daniel Schraivogel, Andreas R. Gschwind, Jennifer H. Milbank, Daniel R. Leonce, Petra Jakob, Lukas Mathur, Jan O. Korb, Christoph A. Merten, Lars Velten, and Lars M. Steinmetz. Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nature Methods*, 17(6):629–635, 2020-06. ISSN 1548-7105. doi: 10.1038/s41592-020-0837-5.
- [307] Teng Fei, Wei Li, Jingyu Peng, Tengfei Xiao, Chen-Hao Chen, Alexander Wu, Jialiang Huang, Chongzhi Zang, X. Shirley Liu, and Myles Brown. Deciphering essential cisomes using genome-wide CRISPR screens. *Proceedings of the National Academy of Sciences*, 116(50):25186–25195, 2019-12-10. doi: 10.1073/pnas.1908155116.
- [308] Li Li, Alejandro P. Ugalde, Colinda L. G. J. Scheele, Sebastian M. Dieter, Remco Nagel, Jin Ma, Abhijeet Pataskar, Gozde Korkmaz, Ran Elkon, Miao-Ping Chien, Li You, Pin-Rui Su, Onno B. Bleijerveld, Maarten Altelaar, Lyubomir Momchev, Zohar Manber, Ruiqi Han, Pieter C. van Breugel, Rui Lopes, Peter ten Dijke, Jacco van Rheenen, and Reuven Agami. A comprehensive enhancer screen identifies TRAM2 as a key and

- novel mediator of YAP oncogenesis. *Genome Biology*, 22(1):54, 2021-01-29. ISSN 1474-760X. doi: 10.1186/s13059-021-02272-8.
- [309] Ivan Krivega and Ann Dean. Enhancer and promoter interactions — long distance calls. *Current opinion in genetics & development*, 22(2):79–85, 2012-04. ISSN 0959-437X. doi: 10.1016/j.gde.2011.11.001.
- [310] Alex Bortvin, Mary Goodheart, Michelle Liao, and David C. Page. Dppa3 / Pgc7 / stella is a maternal factor and is not required for germ cell specification in mice. *BMC Developmental Biology*, 4(1):2, 2004. ISSN 1471213X. doi: 10.1186/1471-213X-4-2.
- [311] Hisayuki Amano, Ken Itakura, Masayoshi Maruyama, Tomoko Ichisaka, Masato Nakagawa, and Shinya Yamanaka. Identification and targeted disruption of the mouse gene encoding ESG1 (PH34/ECAT2/DPPA5). *BMC developmental biology*, 6:11, 2006-02-28. ISSN 1471-213X. doi: 10.1186/1471-213X-6-11.
- [312] Naoki Okashita, Nao Sakashita, Ken Ito, Ayaka Mitsuya, Yoshiaki Suwa, and Yoshiyuki Seki. PRDM14 maintains pluripotency of embryonic stem cells through TET-mediated active DNA demethylation. *Biochemical and Biophysical Research Communications*, 466(1):138–145, 2015-10-09. ISSN 1090-2104. doi: 10.1016/j.bbrc.2015.08.122.
- [313] Franziska K. Geis, Melanie Galla, Dirk Hoffmann, Johannes Kuehle, Daniela Zychlinski, Tobias Maetzig, Juliane W. Schott, Adrian Schwarzer, Christine Goffinet, Stephen P. Goff, and Axel Schambach. Potent and reversible lentiviral vector restriction in murine induced pluripotent stem cells. *Retrovirology*, 14(1):34, 2017-05-31. ISSN 1742-4690. doi: 10.1186/s12977-017-0358-1.
- [314] Laura Magill Sack, Teresa Davoli, Qikai Xu, Mamie Z. Li, and Stephen J. Elledge. Sources of Error in Mammalian Genetic Screens. *G3: Genes/Genomes/Genetics*, 6(9):2781–2790, 2016-07-06. ISSN 2160-1836. doi: 10.1534/g3.116.030973.
- [315] Ye-Ji Sim, Min-Seong Kim, Abeer Nayfeh, Ye-Jin Yun, Su-Jin Kim, Kyung-Tae Park, Chang-Hoon Kim, and Kye-Seong Kim. 2iL Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. *Stem Cell Reports*, 8(5):1312–1328, 2017-04-27. ISSN 2213-6711. doi: 10.1016/j.stemcr.2017.04.001.
- [316] Tetsuya Ishimoto, Kensuke Ninomiya, Ran Inoue, Masato Koike, Yasuo Uchiyama, and Hisashi Mori. Mice Lacking BCAS1, a Novel Myelin-Associated Protein, Display Hypomyelination, Schizophrenia-like Abnormal Behaviors, and Upregulation of Inflammatory Genes in the Brain. *Glia*, 65(5):727–739, 2017-05. ISSN 1098-1136. doi: 10.1002/glia.23129.
- [317] Dung-Fang Lee, Jie Su, Yen-Sin Ang, Xonia Carvajal-Vergara, Sonia Mulero-Navarro, Carlos F. Pereira, Julian Gingold, Hung-Liang Wang, Ruiying Zhao, Ana Sevilla, Henia Darr, Andrew J. K. Williamson, Betty Chang, Xiaohong Niu, Francesca Aguilo, Elsa R. Flores, Yuh-Pyng Sher, Mien-Chie Hung, Anthony D. Whetton, Bruce D. Gelb, Kateri A. Moore, Hans-Willem Snoeck, Avi Ma'ayan, Christoph Schaniel, and Ihor R. Lemischka. Regulation of Embryonic and Induced Pluripotency by Aurora Kinase-p53 Signaling. *Cell stem cell*, 11(2):179–194, 2012-08-03. ISSN 1934-5909. doi: 10.1016/j.stem.2012.05.020.
- [318] Kendall Morgan. Lentiviral CRISPR Libraries Enable Genome-Scale, Knockout Screening. URL <https://blog.addgene.org/lentiviral-crispr-libraries-enable-genome-scale-knockout-screening>.
- [319] Fumitaka Inoue and Nadav Ahituv. Decoding Enhancers Using Massively Parallel Reporter Assays. *Genomics*, 106(3):159–164, 2015-09-01. ISSN 0888-7543. doi: 10.1016/j.ygeno.2015.06.005.
- [320] Atze T. Das, Liliane Tenenbaum, and Ben Berkhout. Tet-On Systems For Doxycycline-inducible Gene Expression. *Current Gene Therapy*, 16(3):156–167, 2016-06. ISSN 1566-5232. doi: 10.2174/1566523216666160524144041.
- [321] Daphne Collias and Chase L. Beisel. CRISPR technologies and the search for the PAM-free nuclease. *Nature Communications*, 12(1):555, 2021-01-22. ISSN 2041-1723. doi: 10.1038/s41467-020-20633-y.
- [322] Cold Spring Harbor. LB Agar. *Cold Spring Harbor Protocols*, 2009(3):pdb.rec11683, 2009-03. ISSN 1940-3402. doi: 10.1101/pdb.rec11683.

- [323] Cold Spring Harbor. 10X PBS. *Cold Spring Harbor Protocols*, 2007(4):pdb.rec10768, 2007-04. ISSN 1940-3402. doi: 10.1101/pdb.rec10768.
- [324] QIAGEN. QIAGEN Supplementary Protocol: Polyacrylamide gel analysis of oligonucleotides, 2002-12.
- [325] F. L. Graham, J. Smiley, W. C. Russell, and R. Nairn. Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5. *Journal of General Virology*, 36(1):59–72, 1977-07-01. ISSN 1465-2099,. doi: 10.1099/0022-1317-36-1-59.
- [326] Aaron R. Quinlan and Ira M. Hall. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, 2010-03. ISSN 1460-2059. doi: 10.1093/bioinformatics/btq033.
- [327] Caroline A. Schneider, Wayne S. Rasband, and Kevin W. Eliceiri. NIH Image to ImageJ: 25 Years of Image Analysis. *Nature Methods*, 9(7):671–675, 2012-07. ISSN 1548-7091. doi: 10.1038/nmeth.2089.
- [328] Fran Supek, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE*, 6(7):e21800, 2011-07-18. ISSN 1932-6203. doi: 10.1371/journal.pone.0021800.
- [329] Cold Spring Harbor. SDS-PAGE Gel. *Cold Spring Harbor Protocols*, 2015(7):pdb.rec087908, 2015-07-01. ISSN 1940-3402, 1559-6095. doi: 10.1101/pdb.rec087908.
- [330] Heyer Lab. DNA Polyacrylamide Gel Electrophoresis, n.d.. URL [http://microbiology.ucdavis.edu/heyer/wordpress/wp-content/uploads/2013/11/DNA\\_PAGE.pdf](http://microbiology.ucdavis.edu/heyer/wordpress/wp-content/uploads/2013/11/DNA_PAGE.pdf).
- [331] Jian Ye, George Coulouris, Irena Zaretskaya, Ioana Cutcutache, Steve Rozen, and Thomas L. Madden. Primer-BLAST: A Tool to Design Target-Specific Primers for Polymerase Chain Reaction. *BMC Bioinformatics*, 13(1):134, 2012-12. ISSN 1471-2105. doi: 10.1186/1471-2105-13-134.
- [332] Manuel Garber. Garber Lab ATAC-seq Protocol, n.d. URL [https://www.encodeproject.org/documents/860c95a8-fdc4-46bd-b9b5-d44242ff20ef/@download/attachment/ATAC\\_garberlab\\_protocol.pdf](https://www.encodeproject.org/documents/860c95a8-fdc4-46bd-b9b5-d44242ff20ef/@download/attachment/ATAC_garberlab_protocol.pdf).





# Appendix

## 7.1 Additional buffers used in this work

### 7.1.1 Gibson assembly mastermix

The protocol for Gibson assembly mastermix was taken from Gibson *et al.*, 2009 [181]. In order to prepare the Gibson assembly mastermix, 5x isothermal (ISO) buffer must first be prepared:

**Table 3:** Requirements for the preparation of 5x ISO buffer

<b>Component</b>	<b>Quantity</b>
1M Tris-HCl, pH 7.5	3 ml
1M MgCl <sub>2</sub>	300 $\mu$ l
NAD <sup>+</sup> (New England BioLabs)	300 $\mu$ l
1M Dithiothreitol (DTT; Carl Roth)	300 $\mu$ l
100 mM dATP (New England BioLabs)	60 $\mu$ l
100 mM dCTP (New England BioLabs)	60 $\mu$ l
100 mM dGTP (New England BioLabs)	60 $\mu$ l
100 mM dTTP (New England BioLabs)	60 $\mu$ l
Polyethylene glycol 8000 (PEG-8000; Sigma-Aldrich)	1.5 g
dH <sub>2</sub> O	<i>ad</i> 6 ml

Prepared 5x ISO buffer was aliquoted and stored at -20°C, prior to use in the Gibson cloning master mix. To prepare the master mix, the following components were mixed:

**Table 4:** Requirements for the preparation of Gibson assembly master mix:

<b>Component</b>	<b>Quantity</b>
40 U/ $\mu$ l Taq ligase (New England BioLabs)	50 $\mu$ l
5x ISO buffer	100 $\mu$ l
1 U/ $\mu$ l T5 Exonuclease (New England BioLabs)	2 $\mu$ l
Phusion polymerase (provided by Department of Developmental Biology, Georg-August University of Göttingen)	6.25 $\mu$ l
dH <sub>2</sub> O	216.75 $\mu$ l

Gibson assembly master mix was aliquoted in 15  $\mu$ l aliquots and stored at -20°C for up to one year. To perform the Gibson method, DNA fragments and dH<sub>2</sub>O *ad* 5  $\mu$ l are combined with 15  $\mu$ l Gibson assembly mastermix and incubated at 50°C for 15-60 minutes, and 1-2  $\mu$ l is then immediately transformed into transformation-competent bacteria.

## 7.1.2 LB Agar

**Table 5:** Requirements for the preparation of LB agar

<b>Component</b>	<b>Quantity</b>
High Gel-Strength Agar (Serva)	12 g
NaCl (AppliChem)	6 g
Tryptone (Oxoid)	6 g
Yeast extract (Oxoid)	3 g
dH <sub>2</sub> O	<i>ad</i> 600 ml

The recipe for LB Agar was taken from the 2009 Cold Spring Harbor (CSH) protocol [322]. The agar was autoclaved, and 100  $\mu\text{g}/\text{ml}$  ampicillin was added to the agar once the glass bottle had reached a temperature that was safe to touch. The agar was immediately poured into 10 cm or 245 x 245 mm square bioassay dishes under a sterile hood until cooled to RT. Plates were then stored upside down and sealed in plastic bags at 4°C until use.

## 7.1.3 PBS, 10x stock

**Table 6:** Requirements for the preparation of 10x PBS

<b>Component</b>	<b>Volume</b>
Na <sub>2</sub> HPO <sub>4</sub> * 2H <sub>2</sub> O (Carl Roth)	17.8 g
NaCl	80 g
KCl (Carl Roth)	2 g
KH <sub>2</sub> PO <sub>4</sub> (AppliChem)	2 g
dH <sub>2</sub> O	<i>ad</i> 1 L

The 10X PBS stock recipe was modified from the 2007 CSH protocol [323]. The stock was autoclaved and diluted 1:10 in dH<sub>2</sub>O prior to use in IF stainings and as a component of PBS-T.

## 7.1.4 TBE, 10x stock

**Table 7:** Requirements for the preparation of 10x TBE buffer

<b>Component</b>	<b>Quantity to be added:</b>
TRIS PUFFERAN <sup>®</sup>	108 g
Boric acid (Sigma-Aldrich)	55 g
0.5 M EDTA (Carl Roth), pH 8.0	40 ml
dH <sub>2</sub> O	<i>ad</i> 1 L

The 10x TBE stock recipe was taken from QIAGEN [324]. TBE 10x stock was filtered through a 0.45  $\mu\text{m}$  filter, and diluted 1:10 in dH<sub>2</sub>O prior to use.

## 7.2 Antibodies

**Table 8:** Antibodies used during the course of this work

Antibody target:	Species of origin:	Conj-ugation:	Dilution:	Company, Catalog No.:
AP2 $\gamma$	Mouse	N/A	1:50 IF 1:200 WB	Santa Cruz Antibodies, sc-12762
Chicken IgG	Goat	N/A	1:100 IF	Invitrogen, A11039
DAZL	Rabbit	N/A	1:100 IF	Abcam, ab34139
GAPDH	Mouse	N/A	1:2000 WB	antibodies.com, ABIN933343
GFP	Chicken	N/A	1:2000 IF	Abcam, ab13970
Goat IgG	Rabbit	CF <sup>TM</sup> -647	1:300 IF	SAB4600188
Mouse IgG	Goat	Alexa-488	1:300 IF	Invitrogen, A11001
Mouse IgG	Goat	Alexa-555	1:300 IF	Invitrogen, A21425
Mouse IgG	Horse	HRP	1:300 WB	Cell Signaling, 7076P2
tdTomato	Goat	N/A	1:100 IF	Sicgen, AB8181-200

In the above table, IgG indicated immunoglobulin G, IF indicates indirect immunofluorescence, WB indicates Western blotting, and HRP indicates horseradish peroxidase.

## 7.3 Cell culture media

### 7.3.1 2i+LIF medium

ESCs were cultured in 2i+LIF medium as described in Silva *et al.*, 2008 [14]. This medium is composed of the defined medium N2B27, to which LIF and the two inhibitors CHIR99021 and PD0325901 are added. Supplement stocks required for the medium were prepared as follows:

**Table 9:** N2B27 stock preparation

<b>Supplement:</b>	<b>Diluent:</b>	<b>Final conc.:</b>
50 mg apo-Transferrin (Sigma-Aldrich)	500 $\mu$ l dH <sub>2</sub> O	100 $\mu$ g/ml
1.6 g Putrescine (Sigma-Aldrich)	10 ml dH <sub>2</sub> O	160 mg/ml
6 mg Progesterone (Sigma-Aldrich)	10 ml 100% EtOH	600 $\mu$ g/ml
6 mg Sodium selenite (Sigma-Aldrich)	10 ml dH <sub>2</sub> O	600 $\mu$ g/ml

To prepare N2B27 medium, N2 supplement is first produced and filtered with a 0.45  $\mu$ m filter in a sterile hood:

**Table 10:** Requirements for the preparation of N2 supplement

<b>Component:</b>	<b>Volume:</b>
DMEM/F12, without glutamine (Gibco)	1 ml
Insulin solution, human (Sigma)	750 $\mu$ l
7.5% Bovine Albumin Fraction V (Gibco)	300 $\mu$ l
apo-Transferrin stock	300 $\mu$ l
Putrescine stock	30 $\mu$ l
Progesterone stock	9.9 $\mu$ l
Sodium selenite stock	3 $\mu$ l

All of the N2 supplement is then added directly to the components of N2B27 medium, and the completed N2B27 medium stored at 4°C:

**Table 11:** Requirements for the preparation of N2B27 medium

<b>Component:</b>	<b>Volume:</b>
Neurobasal <sup>TM</sup> medium (Gibco)	250 ml
DMEM/F12 medium	249 ml
B27 supplement (Gibco)	5 ml
P/S	5 ml
L-glutamine	5 ml
N2 supplement	all

Prior to use with cells, 2i+LIF medium is prepared by directly adding the inhibitors to 37°C pre-warmed N2B27 medium:

**Table 12:** Requirements for the preparation of 2i+LIF medium

<b>Component:</b>	<b>Volume:</b>
N2B27	20 ml
10 $\mu\text{g}/\text{ml}$ LIF	20 $\mu\text{l}$
10 mM CHIR99021	6 $\mu\text{l}$
10 mM PD0325901	2 $\mu\text{l}$

The sterile 2i+LIF medium may be stored at 4°C for up to one week. Both the inhibitors and LIF were provided by the Wellcome-MRC Cambridge Stem Cell Institute.

## 7.4 Cell lines used in this study

In the above table, mESC indicates mouse embryonic stem cells, KO1-3 indicates B1-based AP2 $\gamma$ -knockout line 1, 2, or 3, and S-KO indicates the SGET-based AP2 $\gamma$ -knockout line.

## 7.5 Computer programs and bioinformatics tools

**Table 14:** List of software and bioinformatics tools used in this work

<b>Program or tool:</b>	<b>Produced by:</b>	<b>Version No.:</b>
Affinity Designer	Serif (Europe), Ltd.	1.10.4
annotateMyIDs	Dunning, 2017 [205]	N/A
Bedtools	Quinlan & Hall, 2010 [326]	2.29.1
Bowtie2	Langmead & Salzberg, 2012 [216]	2.3.4.1
Cutadapt	Martin, 2011 [201]	3.7
deepTools	Ramírez <i>et al.</i> , 2016 [213]	3.0.1
FastQC	Andrews, 2015 [199]	0.11.4 (ATAC)
Flowjo	Becton, Dickinson and Company	10.7.1
Galaxy	Afgan <i>et al.</i> , 2018 [198]	22.01.rc1
gplots	Warnes <i>et al.</i> [207]	2.17.0
Homer	Heinz <i>et al.</i> , 2010 [221]	4.8
IGV	Broad Institute; Thorvaldsdóttir <i>et al.</i> , 2013 [212]	2.10.3
ImageJ 1.53K	Schneider <i>et al.</i> , 2012 [327]	1.53p
Image Lab Touch	Bio-Rad	N/A
Macs2	Liu, 2019 [217]	2.1.2
Microsoft <sup>®</sup> Excel	Microsoft	16.58
MultiQC	Ewels <i>et al.</i> , 2016 [200]	1.11
Picard	Broad Institute, 2019 [203]	2.20.2 (ATAC)
Prism 9	GraphPad	9.3.1
R	R Core Team, 2021 [218]	4.1.1
REVIGO	Ruder Bošcović Institute [328]	N/A
SAMtools	Danecek <i>et al.</i> , 2021 [211]	1.9 (ATAC)
SnapGene <sup>®</sup>	Insightful Science	6.0.2
Subread	Liao <i>et al.</i> , 2014 [204]	1.6.3
Trim Galore	Krueger [215]	0.4.1
Zotero	Corporation for Digital Scholarship	5.0.96.4
ZEISS Zen	ZEISS	N/A

## 7.6 Gels

### 7.6.1 SDS-PAGE gels

**Table 15:** Requirements for the preparation of a 10% SDS-PAGE gel

10% separation gel component	Volume	4% stacking gel component	Volume
30% Acrylamide (37.5:1; Carl Roth)	3.3 ml	30% Acrylamide	1.3 ml
dH <sub>2</sub> O	4.15 ml	dH <sub>2</sub> O	6.15 ml
1.5 M Tris-HCl, pH 8.8	2.5 ml	0.5 M Tris-HCl, pH 6.8	2.5 ml
20% SDS (Carl Roth)	50 $\mu$ l	20% SDS	50 $\mu$ l
10% APS ( )	32 $\mu$ l	10% APS	100 $\mu$ l
TEMED (Bio-Rad)	10 $\mu$ l	TEMED	10 $\mu$ l

The recipe for SDS-PAGE gels was modified from the 2015 CSH protocol [329]. All components for the 10% separation gel were mixed, excluding the APS and TEMED. APS and TEMED were then added and the solution was immediately cast. 500  $\mu$ l 100% isopropanol was then gently pipetted on top of the gel to remove bubbles and to prevent the gel from losing moisture. Once the gel had polymerized, all components of the 4% stacking gel excluding APS and TEMED were thoroughly mixed. APS and TEMED were then added, and the solution was mixed and immediately cast onto the stacking gel. 10-well Mini-PROTEAN<sup>®</sup> 1.0 mm combs were then inserted, and once the gel had polymerized it was immediately used for SDS-PAGE.

### 7.6.2 TBE gel, 10%

**Table 16:** Requirements for the preparation of a 10% TBE gel

Component	Volume
30% Acrylamide	4 ml
dH <sub>2</sub> O	5.6 ml
5x TBE	2.4 ml
10% APS	200 $\mu$ l
TEMED	10 $\mu$ l

Protocol for the preparation of a TBE gel was taken from the Wolf-Dietrich Heyer lab protocol [330]. The preparation of TBE buffer is described in Appendix 7.



## 7.7 Materials

### 7.7.1 Cell culture

**Table 17:** Disposables used in cell culture during this work

<b>Product:</b>	<b>Company:</b>	<b>Product No.:</b>
4-well Nunc <sup>TM</sup> Cell-Culture Multiwell Plates, treated	Thermo Scientific	176740
5 mL Polystyrene Round-Bottom Tube with Cell-Strainer Cap	Falcon	352235
6-well Nunc <sup>TM</sup> Cell-Culture Treated Multidishes	Thermo Scientific	140685
10 cm Nunc <sup>TM</sup> EasYDish <sup>TM</sup> Dishes	Thermo Scientific	150464
12-well Nunc <sup>TM</sup> Cell-Culture Multiwell Plates, treated	Thermo Scientific	150628
Acrodisc <sup>®</sup> Syringe Filters with Supor <sup>®</sup> Membrane, Sterile - 0.2 $\mu\text{m}$ , 13 mm	Pall	4602
Cell Scraper S	TPP	99002
Costar <sup>TM</sup> 96-Well Microtiter Plate	Corning	7007
Countess <sup>TM</sup> Cell Counting Chamber Slides	Invitrogen	C10228
EZSPHERE <sup>TM</sup> 6-well microwell plate	Asahi Glass Company	AG4810-900SP
Hecht Karl <sup>TM</sup> Cover Glasses, 18 mm $\varnothing$	Hecht Karl	41001118
Millex-HV 0.45 $\mu\text{M}$ PVDF filter	Merck	SLHVM33RS
Nunc <sup>TM</sup> MicroWell <sup>TM</sup> 96-Well, Nunclon Delta-Treated, Flat-Bottom Microplate	Thermo Scientific	167008

**Table 18:** Media and reagents used in cell culture during this work

<b>Product:</b>	<b>Produced by:</b>	<b>Product No.:</b>
2-Mercaptoethanol (50 mM)	Gibco	31350010
Activin A	Wellcome-MRC Cambridge Stem Cell Institute	N/A
apo-Transferrin	Sigma-Aldrich	T1147-500MG
B27 <sup>TM</sup> Supplement (50x), serum-free	Gibco	17504044
Bovine Albumin Fraction V (7.5% solution)	Gibco	15260037
CHIR99021	Wellcome-MRC Cambridge Stem Cell Institute	N/A
Dimethylsulfoxide (DMSO)	Sigma-Aldrich	D8418-100ML
DMEM/F12, without glutamine	Gibco	21331020
Fetal Bovine Serum, qualified, Brazilian origin (FBS)	Gibco	10270106
Gelatin	Merck	4070
Glasgow's MEM (GMEM)	Gibco	21710025
GlutaMAX <sup>TM</sup> supplement	Gibco	35050061
Human Plasma Fibronectin (purified protein)	Sigma-Aldrich	FC010-10MG
Insulin Solution (Human)	Sigma-Aldrich	I9278-5ML
KnockOut <sup>TM</sup> Serum Replacement (KSR)	Gibco	10828010
L-Glutamine (200 mM)	Gibco	25030081
Lipofectamine <sup>TM</sup> 2000 Transfection Reagent	Invitrogen	11668030
MEM Non-essential amino acids solution (NEAA), 100x	Gibco	11140-050
Mouse Leukemia Inhibitory Factor (m-LIF)	Wellcome-MRC Cambridge Stem Cell Institute	0717-010
Neurobasal <sup>TM</sup> Medium	Gibco	21103049
Opti-MEM <sup>TM</sup> I 1x reduced-serum medium	Gibco	31985062
PBS pH 7.2 (1x)	Gibco	20012-019
PD0325901	Wellcome-MRC Cambridge Stem Cell Institute	N/A

<b>Product:</b>	<b>Produced by:</b>	<b>Product No.:</b>
Penicillin-Streptomycin (10,000 U/mL)	Gibco	15140122
PLUS <sup>TM</sup> Reagent	Invitrogen	11514015
Progesterone	Sigma-Aldrich	P8783-1G
Puromycin Dihydrochloride	Gibco	A11138-03
Putrescine dihydrochloride	Sigma-Aldrich	P5780-5G
Sodium pyruvate solution, 100 mM	Sigma-Aldrich	S8636-100ml
Sodium selenite	Sigma-Aldrich	S5261-10G
TrypLE <sup>TM</sup> Express Enzyme (1x), without Phenol-red	Gibco	12604013
UltraPure <sup>TM</sup> Distilled Water DNase/RNase Free	Invitrogen	10977-035

## 7.7.2 Chemicals and reagents

**Table 19:** Non-cell culture chemicals and reagents used during this work

<b>Product:</b>	<b>Produced by:</b>	<b>Product No.:</b>
1 kb DNA Ladder	New England BioLabs	N3232L
1,4-Dithiothreitol (DTT)	Carl Roth	6908.2
4x Laemmli Sample Buffer	Bio-Rad Laboratories	1610747
100 bp DNA ladder	New England BioLabs	N3231L
Agarose	Th. Geyer	9920-500G
Agencourt <sup>®</sup> RNAClean <sup>®</sup> XP Beads	Beckman Coulter	A66514
Albumin fraction V	Carl Roth	8076.2
Ammonium persulfate	Sigma-Aldrich	A3678-25G
Ampicillin	Serva	13397
Automatic Setup Beads	Sony	LE-B3001
Boric acid	Sigma-Aldrich	B7660
DAPI	Roche Diagnostics	15733122
Difco <sup>™</sup> Skim Milk	BD	232100
Disodium phosphate dihydrate (Na <sub>2</sub> HPO <sub>4</sub> * 2H <sub>2</sub> O)	Carl Roth	4984.1
dNTPs	Invitrogen	46-0122
DRAQ5 <sup>™</sup> Fluorescent Probe Solution (5 mM)	Invitrogen	62251
D(+)-Saccharose	Carl Roth	4621.1
Ethanol absolute	VWR	20821.321
EDTA	Carl Roth	8043.2
Glycine	Carl Roth	3187.5
GlycoBlue <sup>™</sup> Co-precipitant	Invitrogen	AM9515
HEPES PUFFERAN <sup>™</sup>	Carl Roth	9105.4
High Gel-Strength Agar	Serva	11396
Hydrochloric acid (HCl) 37%	Carl Roth	4625.2
Igepal CA-630	Sigma-Aldrich	I-3021
Methanol	VWR	20864.32

<b>Product:</b>	<b>Produced by:</b>	<b>Product No.:</b>
Magnesium chloride hexahydrate (MgCl <sub>2</sub> * 6H <sub>2</sub> O)	Carl Roth	2189.1
NAD+	New England BioLabs	B9007S
NEBuffer 4	New England BioLabs	B7004S
OCT Embedding Matrix	Carl Roth	6478.2
PageRuler <sup>®</sup> Prestained Protein Ladder	Thermo Scientific	26616
Pierce <sup>™</sup> ECL Western Blotting Solution	Thermo Scientific	32209
Paraformaldehyde 16% Solution	Agar Scientific	R1026
Polyethylene glycol 8000 (PEG 8000)	Sigma-Aldrich	20,245-2
Potassium chloride (KCl)	Carl Roth	6781.1
Potassium phosphate (KH <sub>2</sub> PO <sub>4</sub> )	AppliChem	A1043,1000
Rotiphorese <sup>®</sup> Gel 30 (37.5:1) acrylamide solution	Carl Roth	3029.1
Sodium acetate trihydrate	Alfa Aesar	S/2040/53?
Sodium chloride (NaCl)	AppliChem	A3597,1000
Sodium dodecyl sulfate (SDS)	Carl Roth	2326.2
SYBR <sup>™</sup> Green I Nucleic Acid Stain	Lonza	50513
SYBR <sup>™</sup> Safe DNA Gel Stain	Invitrogen	S33102
TEMED	Bio-Rad Laboratories	161-0800
TRIS PUFFERAN <sup>®</sup>	Carl Roth	AE15.2
Triton <sup>®</sup> X-100	Fisher Bioreagents	BP151-100
Tryptone	Oxoid	LP0042
Tween <sup>®</sup> 20	Fisher Bioreagents	BP337-100
UltraPure <sup>™</sup> Distilled Water DNase/RNase Free	Invitrogen	10977-035
VECTASHIELD <sup>®</sup> PLUS Antifade Mounting Medium	Vector Laboratories, Inc.	H-1900
Water (for RNA Work)	Fisher Bioreagents	BP561-1
Yeast Extract	Oxoid	LP0021

### 7.7.3 Disposables

**Table 20:** Non-cell culture disposables used during this work

<b>Product:</b>	<b>Produced by:</b>	<b>Product No.:</b>
Amersham™ Protran™ 0.45 μm NC Nitrocellulose Blotting Membrane	Cytiva	10600093
Corning® 245 mm square bioassay dish	Sigma-Aldrich	CLS431111
Cover slips 24 x 40 mm#1	Menzel Gläser	MENZBB024040M113
Disposable Base Molds, 24 x 24 x 5mm	Epredia	41740
Electroporation Cuvette Pre-Sterilized, 1 mm Gap	Molecular BioProducts	5510
Microscopic Slides ground 90° frosted	Menzel Gläser	AGAA000001##12E
Microseal® 'B' Seals	Bio-Rad Laboratories	MSB1001
Multiplate® PCR Plates™ 96-well, clear	Bio-Rad Laboratories	MLL9601
Sorting Chip	Sony	LE-C3210
Superfrost™ Plus Adhesion Microscope Slides	Epredia	J1800AMNZ

## 7.7.4 Enzymes and mastermixes

**Table 21:** Enzymes and mastermixes used during this work

Product	Produced by	Product No.
BbsI-HF <sup>®</sup>	New England BioLabs	R3539S
BamHI-HF <sup>®</sup>	New England BioLabs	R3136S
KAPA SYBR <sup>®</sup> FAST qPCR Master Mix (2X) Universal	Kapa Biosystems	KK4618
LunaScript <sup>™</sup> RT SuperMix Kit	New England BioLabs	E3010G
NEB <sup>®</sup> PCR Cloning Kit (without competent cells)	New England BioLabs	E1203S
NEBNext <sup>®</sup> High-Fidelity 2X PCR Master Mix	New England BioLabs	M0541S
NheI-HF <sup>®</sup>	New England BioLabs	R3131S
OneTaq <sup>®</sup> DNA Polymerase	New England BioLabs	M0480L
Phusion polymerase	Department of Developmental Biology, Georg-August University of Göttingen	N/A
Q5 <sup>®</sup> High-Fidelity DNA Polymerase	New England BioLabs	M0491S
Sall-HF <sup>®</sup>	New England BioLabs	R3138S
SmaI	New England BioLabs	R0141S
T4 DNA Ligase	New England BioLabs	M0202
T5 Exonuclease	New England BioLabs	M0363S
Taq DNA Ligase	New England BioLabs	M0208L



<b>Product</b>	<b>Produced by</b>	<b>Product No.</b>
XbaI	New England BioLabs	R0145S
XhoI	New England BioLabs	R0146S

## 7.7.5 Equipment

**Table 22:** Equipment used during this work

<b>Product</b>	<b>Produced by</b>	<b>Product No.</b>
CFX96 <sup>®</sup> Real-Time PCR Detection System	Bio-Rad Laboratories	185-5096
ChemiDoc <sup>™</sup> Imaging System	Bio-Rad Laboratories	12003153
Colirollers Plating Beads	EMD Millipore Corp	71013-1PKG
Countess II Automated Cell Counter	Invitrogen	AMQAX1000
Dual LED Blue/White Light Transilluminator	Invitrogen	S37103
Gene Pulser XCell Total Electroporation System	Bio-Rad Laboratories	1652660
HERAcell <sup>™</sup> VIOS 160i	Thermo Scientific	50145502
ImmEdge Hydrophobic Barrier (PAP) Pen	Vector Laboratories	H-4000
Leica CM1950	Leica	1491950C4US
Mini-PROTEAN <sup>®</sup> Comb, 10-well, 1.0 mm, 44 $\mu$ l	Bio-Rad	1653359
Mini-PROTEAN <sup>®</sup> Short Plates	Bio-Rad	1653308
Mini-PROTEAN <sup>®</sup> Tetra cell	Bio-Rad	1658000
NanoDrop 2000c Spectrophotometer	Thermo Scientific	ND2000CLAPTOP
SH800S Cell Sorter	Sony	SH800S
Qubit <sup>®</sup> 2.0 Fluorometer	Invitrogen	Q32866
ZEISS AXIO Observer.Z1	ZEISS	N/A
ZEISS LSM 980	ZEISS	N/A

## 7.7.6 Kits

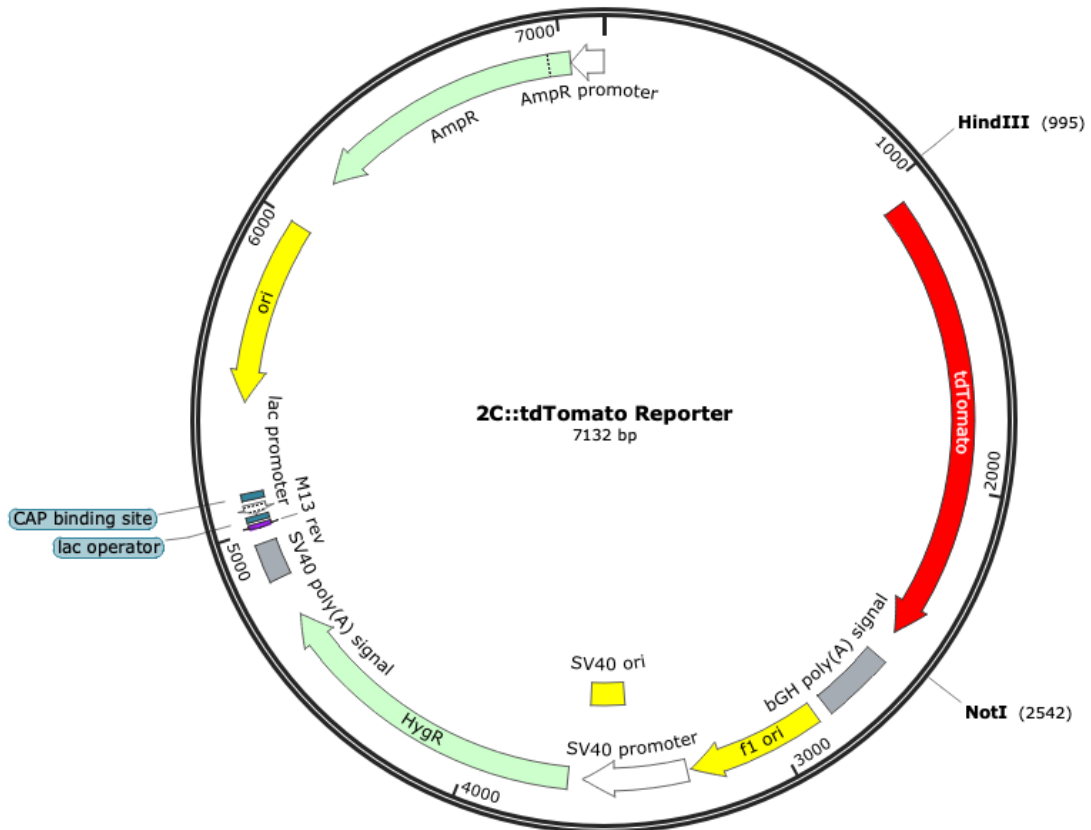
**Table 23:** Kits used during this work

<b>Product</b>	<b>Produced by</b>	<b>Product No.</b>
DNeasy <sup>®</sup> Blood & Tissue Kit (50)	QIAGEN	69504
EndoFree <sup>®</sup> Plasmid Maxi Kit (10)	QIAGEN	12362
MinElute Gel Extraction Kit (50)	QIAGEN	28604
MinElute PCR Purification Kit (50)	QIAGEN	28004
NEBNext <sup>®</sup> Ultra <sup>™</sup> II Directional RNA Library Prep Kit for Illumina <sup>®</sup>	New England BioLabs	E7760S
NEBNext <sup>®</sup> Ultra <sup>™</sup> II DNA Library Prep Kit for Illumina <sup>®</sup>	New England BioLabs	E7645S
Nextera DNA Library Prep Kit	Illumina	FC-121-1030
NucleoSpin <sup>®</sup> Gel and PCR Clean-up	Macherey-Nagel	740609.250
NucleoSpin <sup>®</sup> Plasmid	Macherey-Nagel	740588.250
NucleoSpin <sup>®</sup> Plasmid Transfection-grade Kit	Macherey-Nagel	740490.250
QuantiTect <sup>®</sup> Reverse Transcription Kit (200)	QIAGEN	205313
Qubit <sup>®</sup> RNA HS Assay Kit	Invitrogen	Q32852
Qubit <sup>®</sup> dsDNA HS Assay Kit	Invitrogen	Q32851
Qubit <sup>®</sup> Protein Assay Kit	Invitrogen	Q33211
<i>Quick</i> -DNA <sup>™</sup> Midiprep Plus Kit	Zymo Research	D4075
RNeasy <sup>®</sup> Plus Micro Kit (50)	QIAGEN	73034

## 7.8 Plasmids

### 7.8.1 2C::tdTomato Reporter

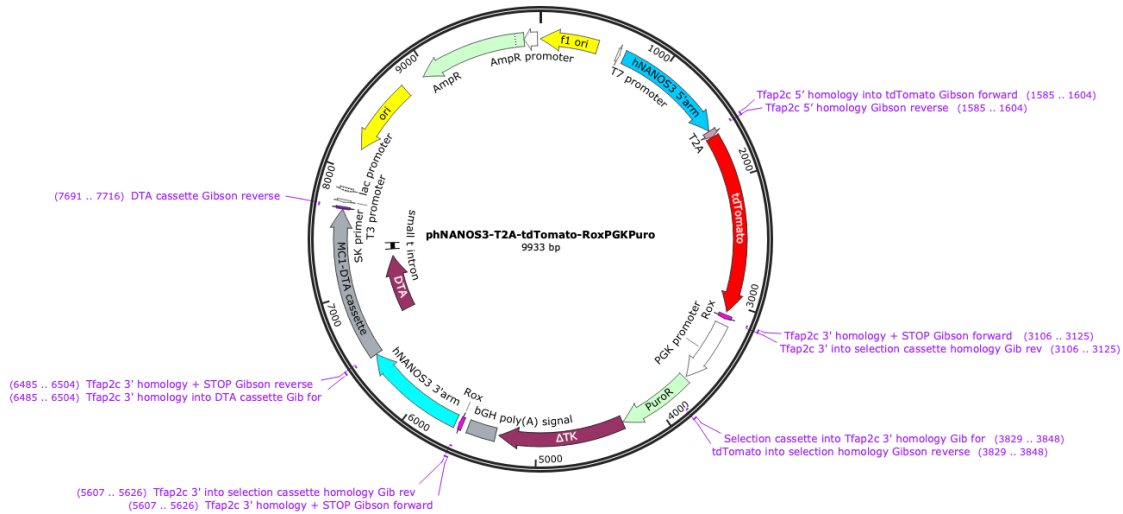
The plasmid 2C::tdTomato Reporter was a gift from Samuel Pfaff [180]. This plasmid encodes the *Tdtomato* reporter gene, as well as hygromycin and ampicillin resistance cassettes. This plasmid is published in MacFarlan *et al.*, 2012 [179].



**Figure 41:** Map of 2C::tdTomato reporter plasmid. All features are noted, as well as RE sites used during the course of this work to linearize the plasmid.

## 7.8.2 pHNANOS3-T2A-tdTomato-RoxPGKPuro

The plasmid pHNANOS3-T2A-tdTomato-RoxPGKPuro was provided by the Azim Surani lab, and was originally designed and cloned by Toshihiro Kobayashi. This plasmid was designed to allow for the insertion of the fluorescent tdTomato reporter into the human *NANOS3* gene using CRISPR/Cas9 and HDR, and thus includes the 5' and 3' homologous recombination arms. The plasmid also features a puromycin resistance cassette, a  $\Delta$ TK fragment, and an MC1-DTA cassette. This plasmid was published by Kobayashi *et al.*, 2017 [175].



**Figure 42:** Map of pHNANOS3-T2A-tdTomato-RoxPGKPuro plasmid. All features are indicated and primers which were used during the course of this study to amplify features of the plasmid are indicated.

### 7.8.3 pCAGGS-Dre-IH

The plasmid pCAGGS-Dre-IH was provided by the Azim Surani lab. This plasmid encodes the Dre recombinase enzyme, and additionally expresses a hygromycin resistance cassette which is internally processed from the Dre recombinase with the aid of an IRES site. This plasmid was published by Kobayashi *et al.*, 2017, where it was called “pCAG-DRE-IH” [175].

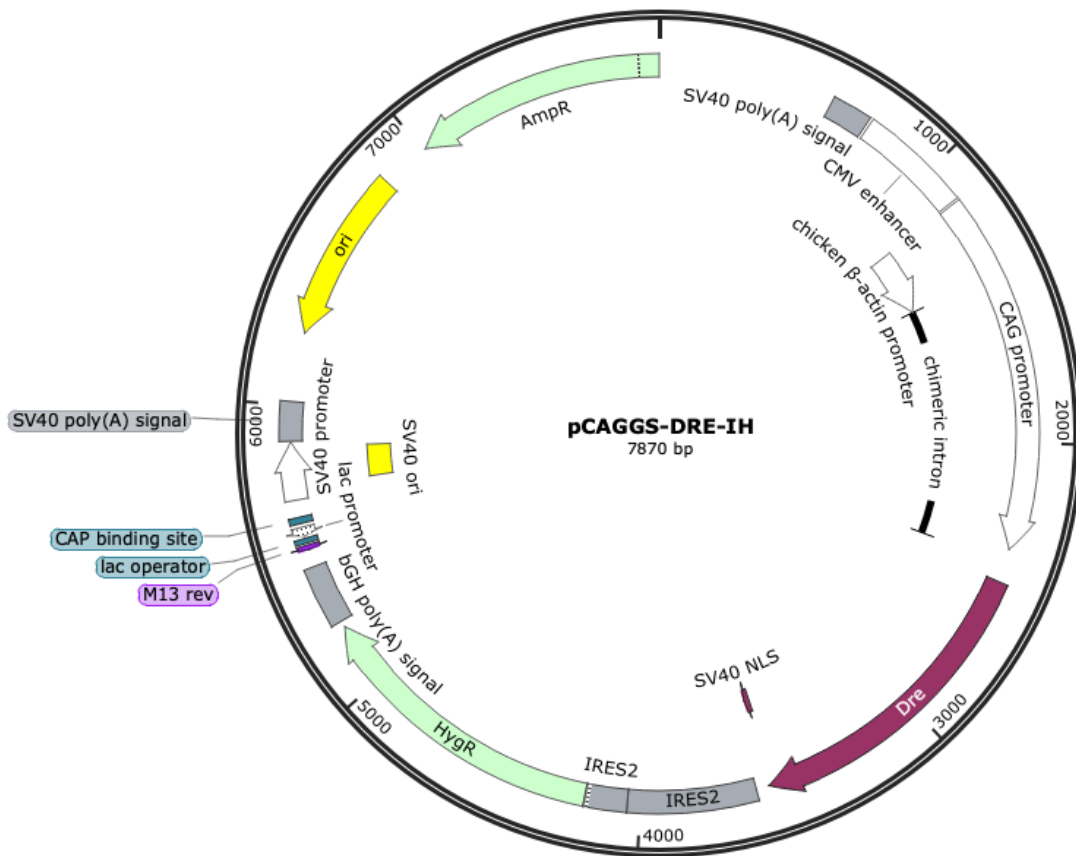
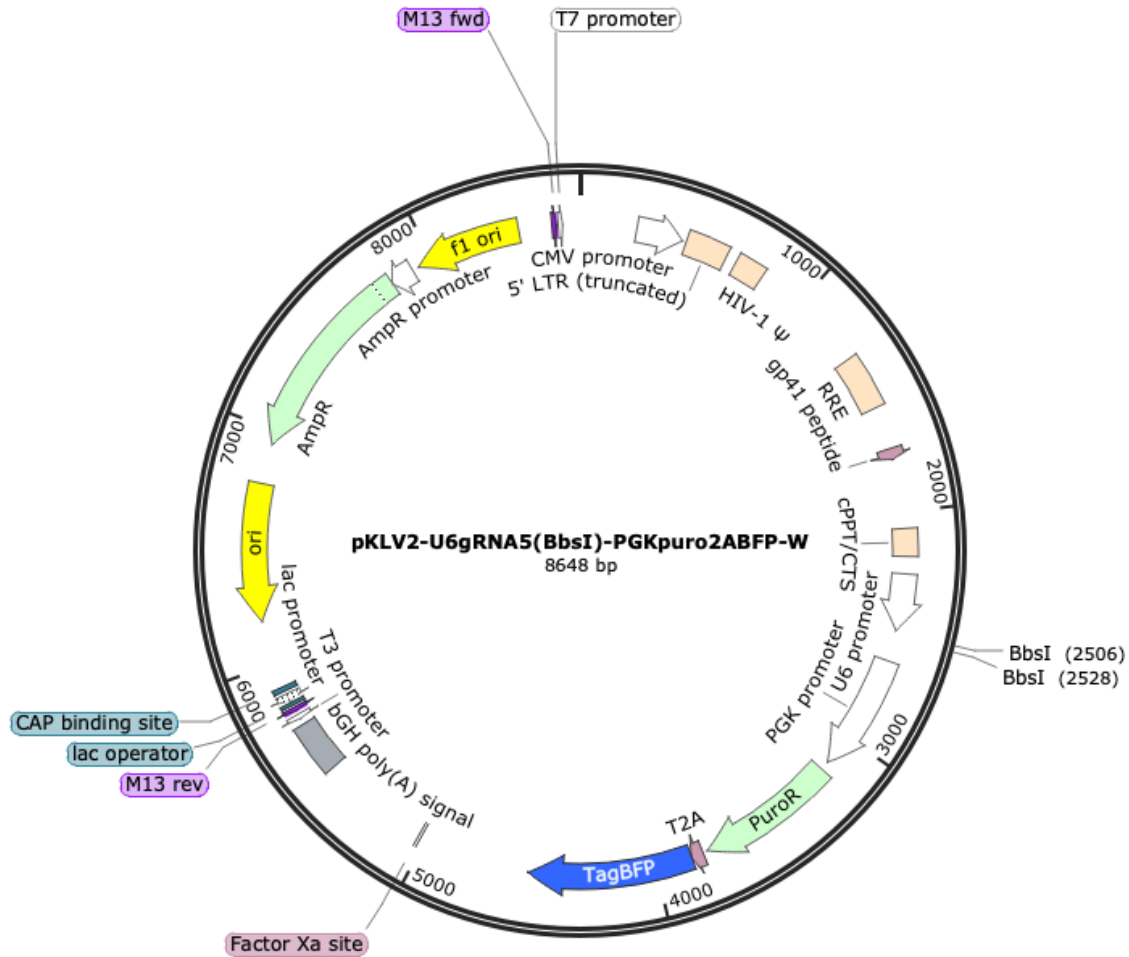


Figure 43: Map of pCAGGS-Dre-IH plasmid. All features are indicated.

### 7.8.4 pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W

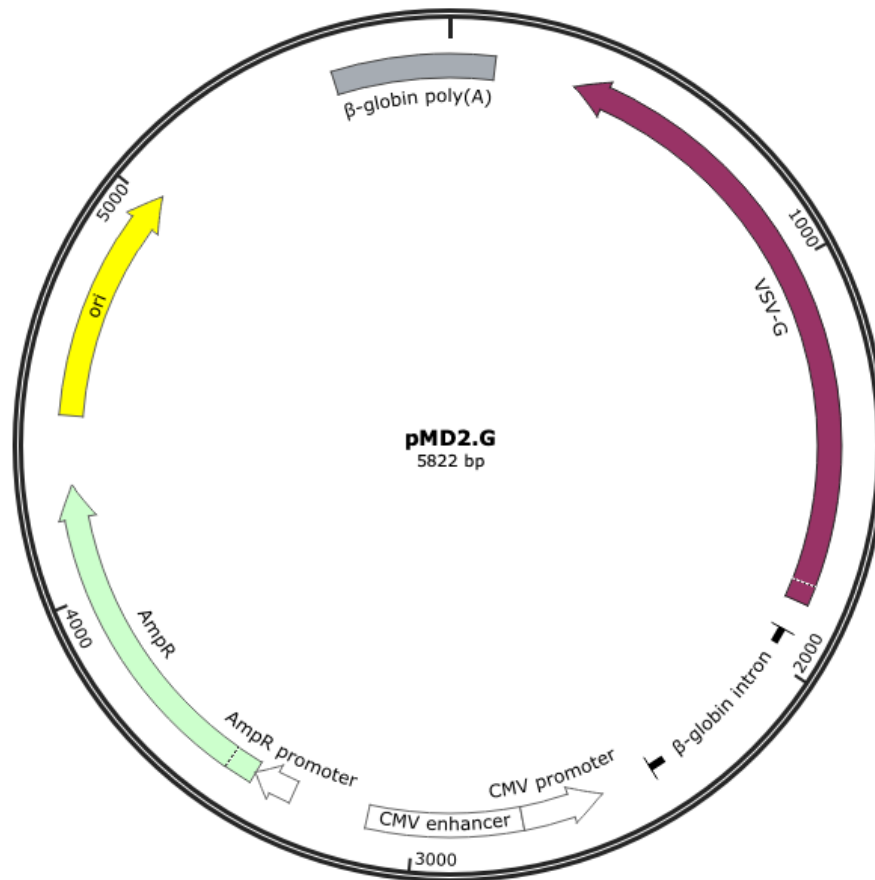
The plasmid pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W (pKLV2-W) was a gift from Kosuke Yusa [227]. This plasmid was RE-digested with BbsI-HF<sup>®</sup> for the insertion of *Tfap2c* tiling screen gRNA library gRNAs. The plasmid is also one of three plasmids, including psPAX2 and pMD2.G, used to produce gRNA-containing lentivirus [156].



**Figure 44:** Map of pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W plasmid. All features and RE-digest sites used in this work are indicated.

### 7.8.5 pMD2.G

The plasmid pMD2.G was a gift from Didier Trono [237], and is unpublished. The plasmid is also one of three plasmids, including psPAX2 and pKLV2-W, used to produce gRNA-containing lentivirus [156].



**Figure 45:** Map of pMD2.G plasmid. All features are indicated.





## 7.8.10 psPAX2

The plasmid psPAX2 was a gift from Didier Trono [238], and is unpublished. The plasmid is also one of three plasmids, including pMD2.G and pKLV2-W, used to produce gRNA-containing lentivirus [156].

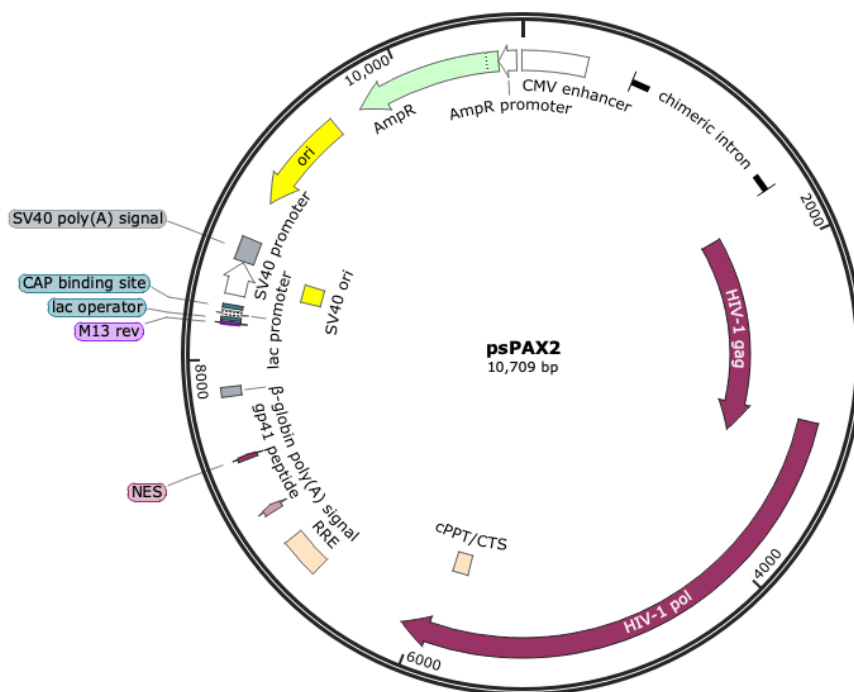
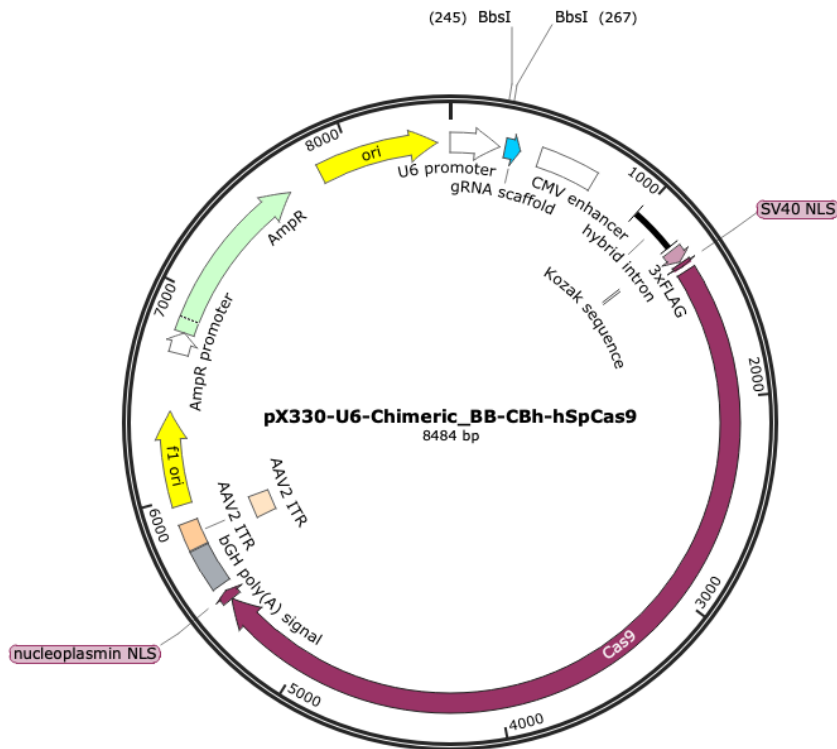


Table 24: Map of psPAX2 plasmid. All features are indicated.

### 7.8.11 pX330-U6-Chimeric\_BB-CBh-hSpCas9

The pX330-U6-Chimeric\_BB-CBh-hSpCas9 (pX330) plasmid was a gift from Feng Zhang [168], and contains the sequence encoding the Cas9 enzyme tagged with a nuclear-localization signal (NLS). The plasmid also contains two BbsI RE cut-sites to allow for gRNA insertion, downstream of the U6 promoter. This plasmid was originally published by Cong *et al.*, 2013 [167].



**Figure 47:** Map of pX330-U6-Chimeric\_BB-CBh-hSpCas9 plasmid. All features are indicated, as well as the BbsI cut-sites relevant to this work.

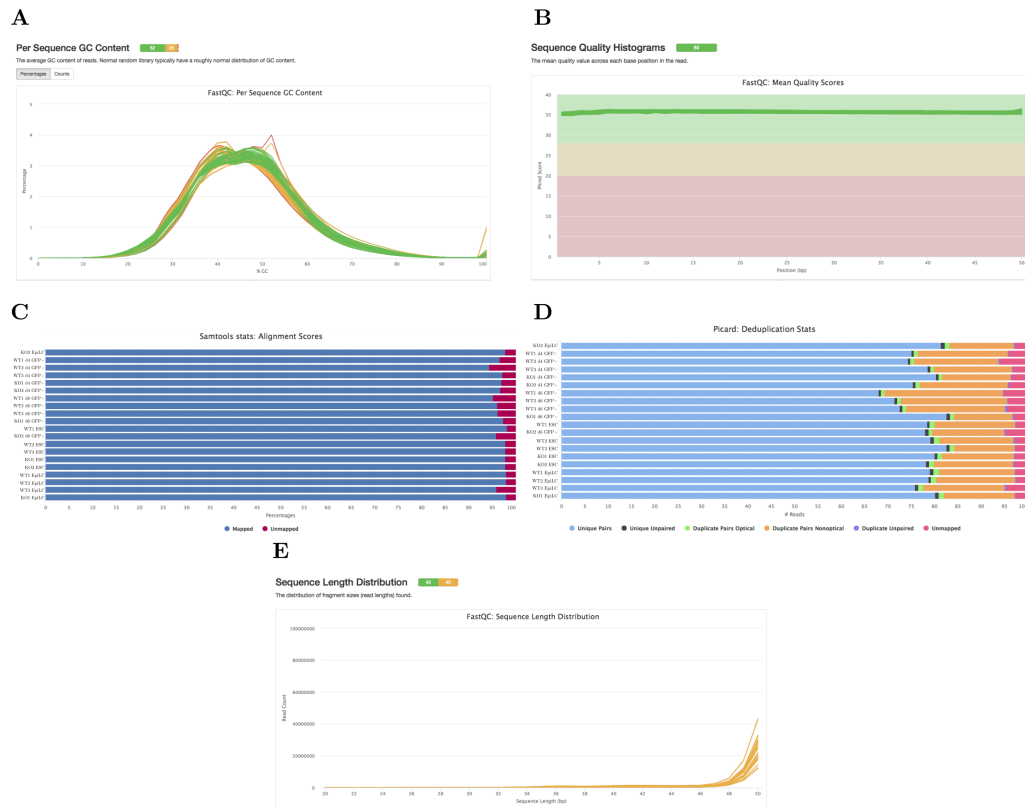
Ultimately, five pX330-based plasmids containing unique gRNAs were generated during the course of this work:

- pX330-a'mTfap2c-exon2-g2
- pX330-a'mTfap2c-exon2-g3
- pX330-a'mTfap2c-stop-g5

## 7.9 Sequencing quality control and supplementary data

### 7.9.1 ATAC-seq

#### 7.9.1.1 ATAC-seq qualitative library assessment

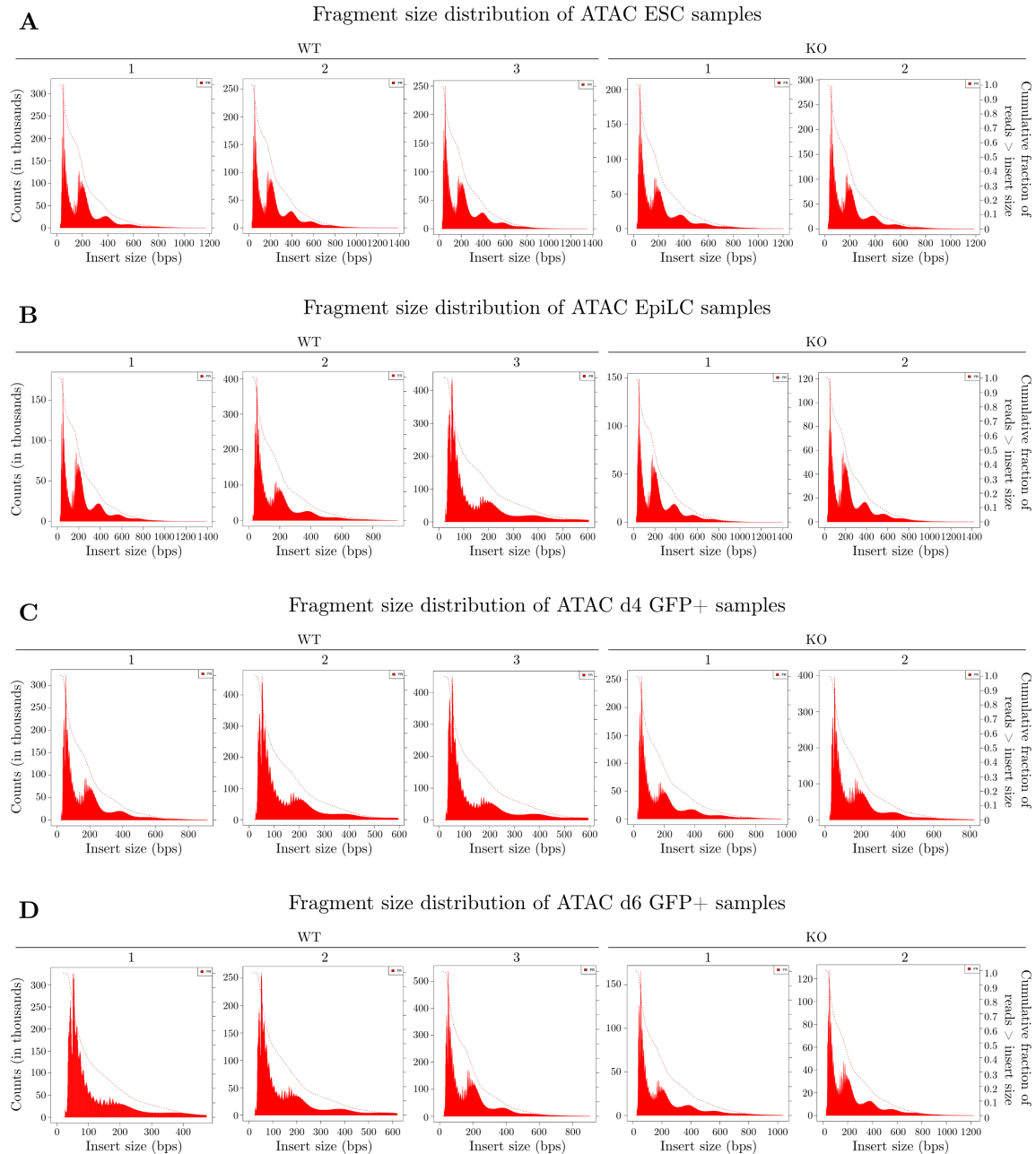


**Figure 48: Quality control of ATAC-seq libraries.** All libraries were aggregated per test with MultiQC. A) FastQC analysis: percentage of GC content of all ATAC-seq libraries pre- and post-Cutadapt-based adapter trimming. All but two samples “passed” as indicated by green or yellow curves; two samples “failed” as indicated by red curves. B) analysis: mean quality scores of all ATAC-seq libraries pre- and post-Cutadapt-based adapter trimming. All samples “passed” as indicated by the green line. C) SAMtools stats of HISAT2 alignment of all ATAC-seq libraries post-Cutadapt adapter-trimming. Blue indicates mapped, red indicates unmapped. D) Picard deduplication statistics of all ATAC-seq libraries post-Cutadapt adapter-trimming. Duplication status is indicated by color, and explained in the associated key. E) FastQC analysis: sequence length distribution of all ATAC-seq libraries pre- and post-Cutadapt-based adapter trimming. All pre-trimming samples passed, as indicated by the yellow line, and all post-trimming samples failed, as indicated by the green line.

Qualitative assessment of the ATAC-seq libraries was performed by testing the libraries for sequence length, GC content, sequence quality, and percentages of genomic alignment and duplication. Two libraries failed the per sequence GC content FastQC test, however behaved otherwise normally. All libraries had between 65-85% unique reads, and >93% genomic alignment. Libraries had high mean quality scores, and behaved

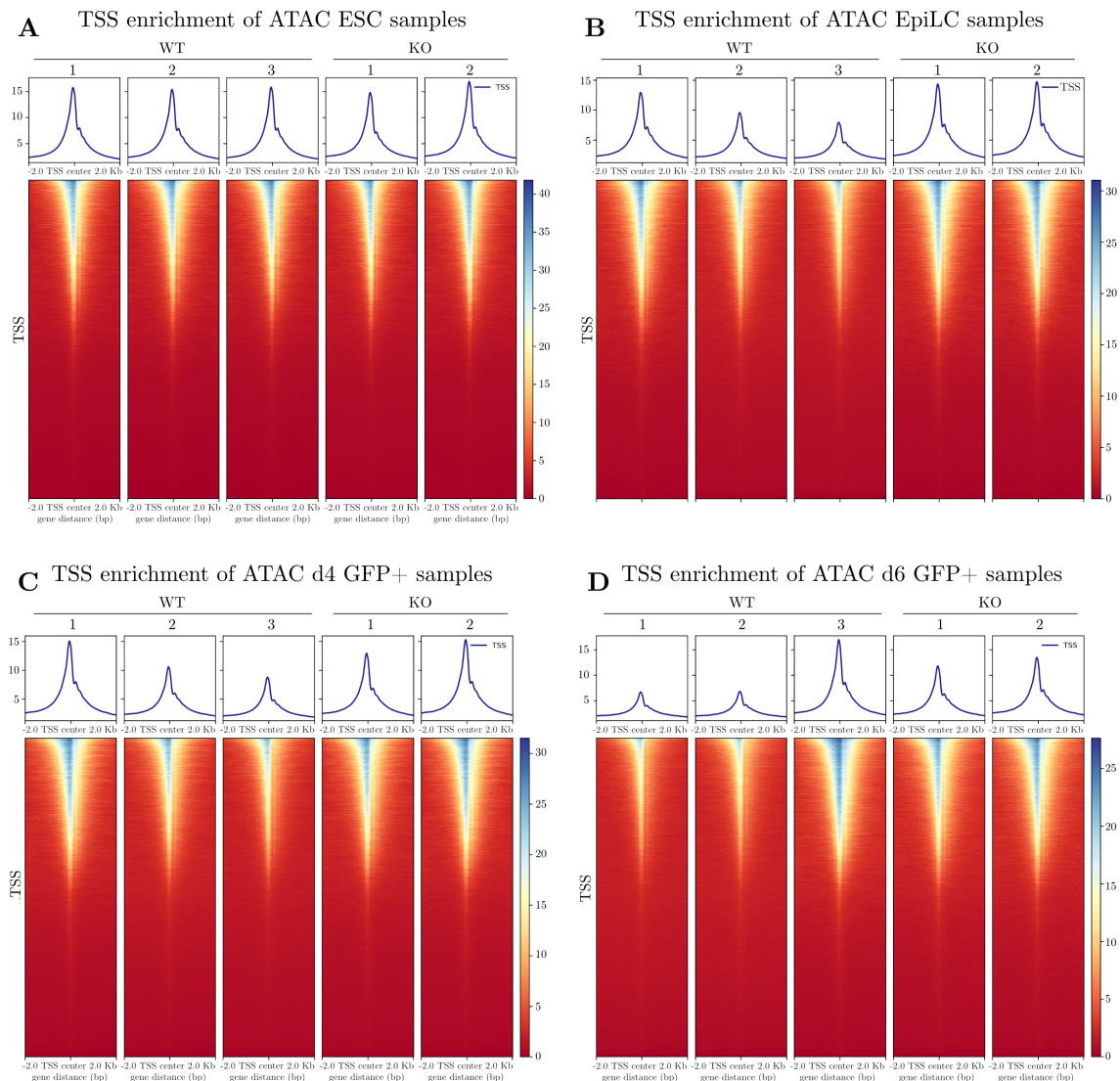
as expected pre- and post-trimming. Based on these assessments, it was decided that the libraries were suitable for use.

### 7.9.1.2 ATAC-seq library fragment length distribution



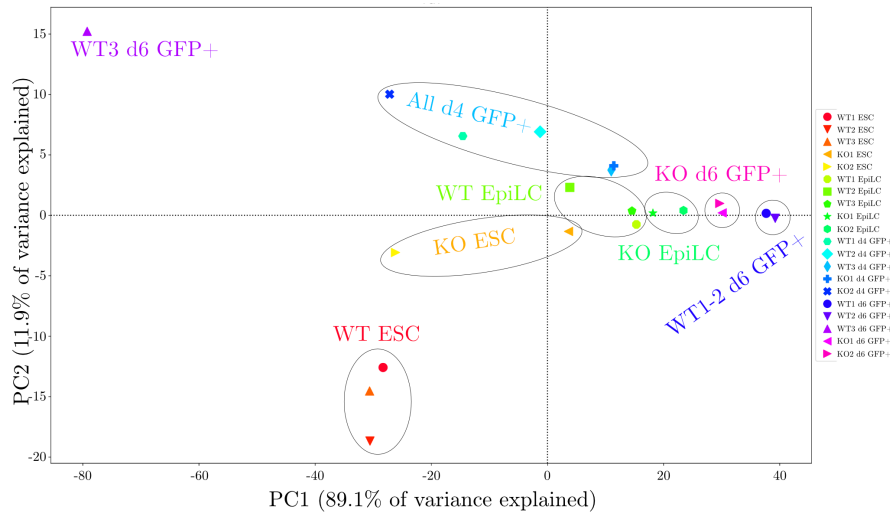
**Figure 49: Fragment length distribution of each ATAC-seq libraries used in this work.** Insert size in bps and counts in thousands are indicated. A) ESC WT and KO ATAC-seq libraries. B) EpiLC WT and KO ATAC-seq libraries. C) d4 GFP+ WT and KO ATAC-seq libraries. D) d6 GFP+ WT and KO ATAC-seq libraries.

### 7.9.1.3 ATAC-seq library transcription start sites



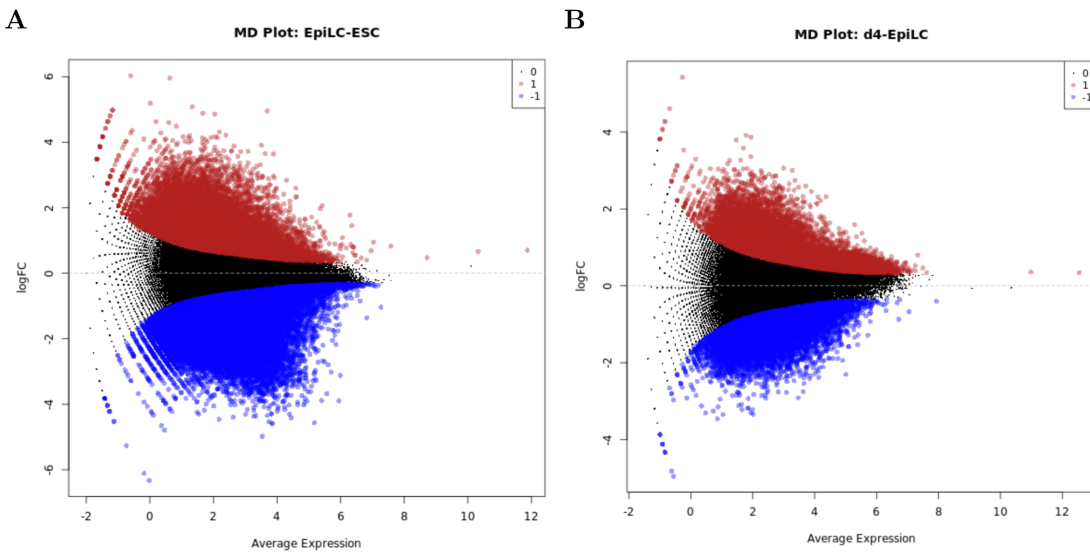
**Figure 50: Transcription start sites of ATAC-seq libraries.** The alignment of mapped reads at transcription start sites (TSS) of each ATAC-seq library is shown, with cell type and KO status indicated, as well as number of reads in thousands (0-15), and distance from TSS in bp are indicated. Color indicates number of reads aligning to the TSS, with blue indicating high alignment and red indicating poor alignment. The average TSS profile per library is shown.

### 7.9.1.4 ATAC-seq PCA



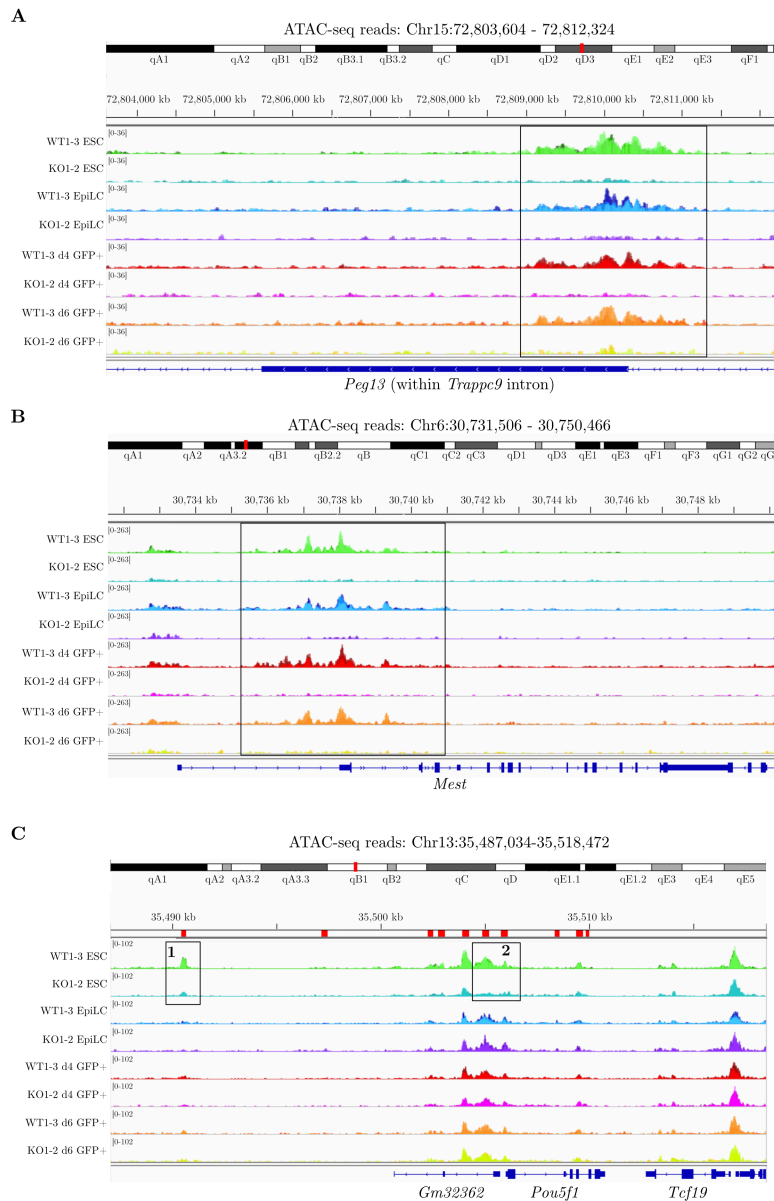
**Figure 51: PCA of ATAC-seq libraries used in this work.** Percentage of variance in PC1 and PC2 are indicated. Clusters by cell type and/or KO status are indicated by black rings, key is also indicated.

### 7.9.1.5 Differentially accessible peaks between WT cell types



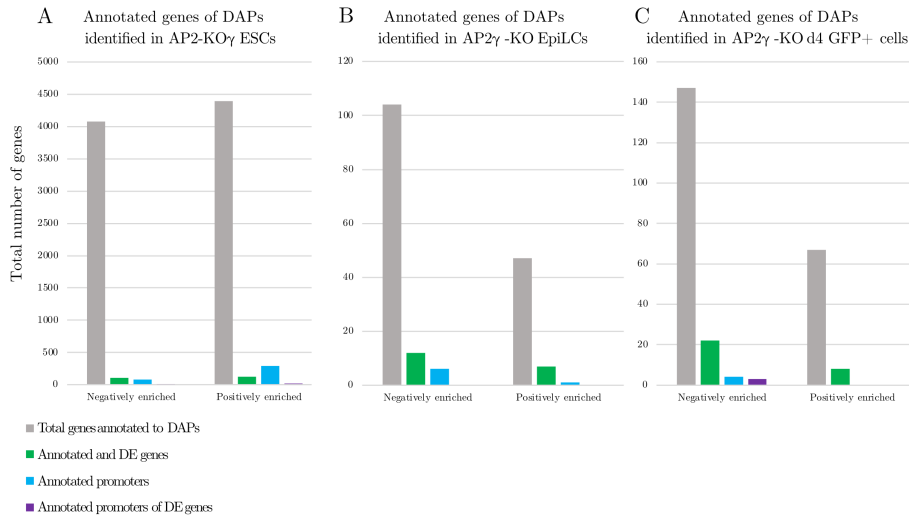
**Figure 52: MD plots of DAPs identified between WT cell-state transitions.** Peaks significantly negatively or positively enriched are indicated, as is the average expression and logFC. A) DAPs enriched in WT EpiLCs relative wt WT ESCs. B) DAPs enriched in WT d4 GFP+ PGCLCs relative to WT EpiLCs.

### 7.9.1.6 Example DAPs of interest identified in ATAC-seq libraries



**Figure 53: IGV tracks demonstrating DAPs identified in genes of interest.** IGV tracks demonstrate WT and AP2 $\gamma$ -KO overlaid Bigwig files representing accumulated reads. Chromosomal coordinates and increments are indicated. A) *Peg13* locus demonstrating DAP in indicated black box. B) *Mest* locus demonstrating DAP in indicated black box. C) *Pou5f1* locus demonstrating two DAPs in indicated black boxes. CREs identified by ENCODE are indicated in red adjacent to the chromosomal tracks.

### 7.9.1.7 Genes annotated to DAPs

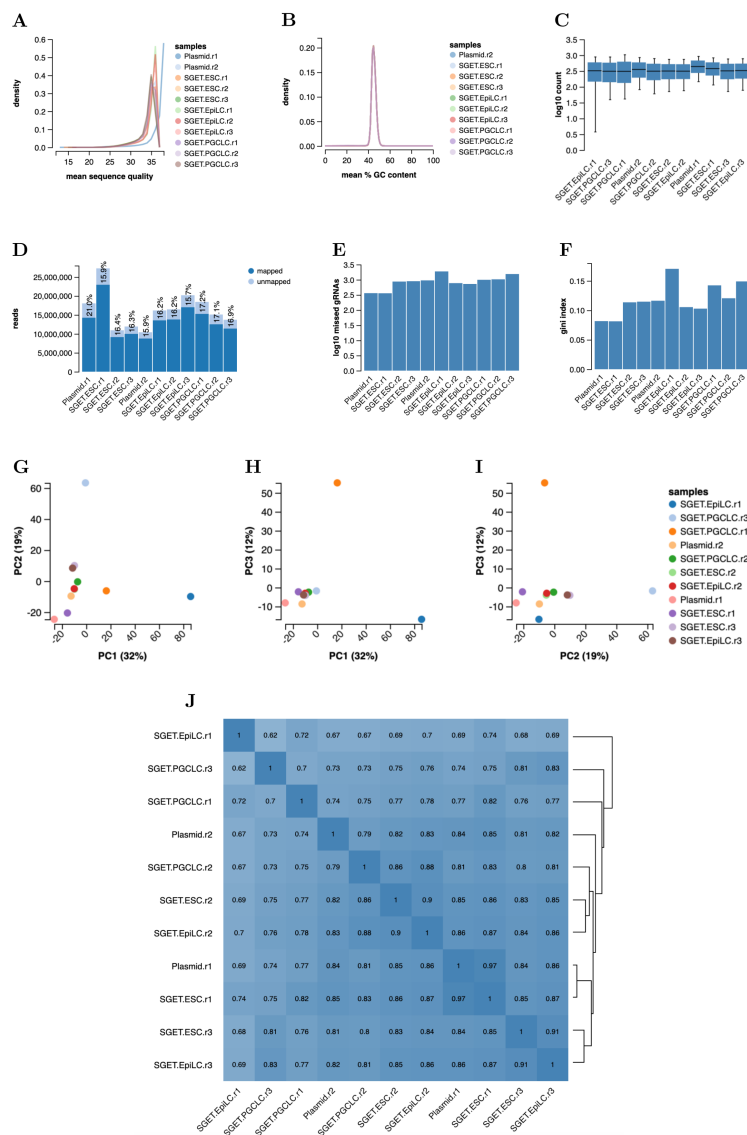


**Figure 54: Graphs indicating number of genes annotated to DAPs in ATAC-seq data.** The total number of genes annotated to DAPs (in grey), genes annotated to DAPs corresponding to DE genes in RNA-seq data (in green), genes annotated to DAPs corresponding to promoter regions (in blue), and genes annotated to DAPs corresponding to promoter regions of genes identified as DE in RNA-seq data (in purple) are shown by cell type and AP2 $\gamma$ -KO status. A) Genes annotated to DAPs in AP2 $\gamma$ -KO ESCs. B) Genes annotated to DAPs in AP2 $\gamma$ -KO EpiLCs. C) Genes annotated to DAPs in AP2 $\gamma$ -KO d4 GFP<sup>+</sup> cells. D) Genes annotated to DAPs in AP2 $\gamma$ -KO d6 GFP<sup>+</sup> cells.



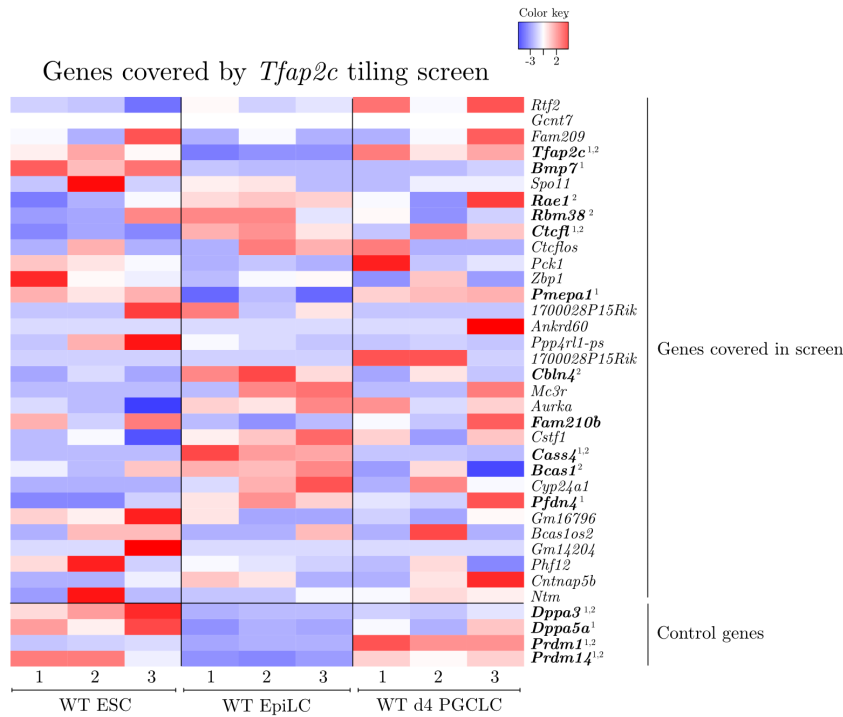
## 7.9.2 *Tfap2c* tiling screen sequencing

### 7.9.2.1 Quality control statistics of *Tfap2c* tiling screen



**Figure 55: *Tfap2c* tiling screen sequencing statistics and sample similarity information.** All plots were generated in MaGECK-VISPR, for all SGET ESC, EpiLC, and PGCLC libraries. A) Mean sequence quality of tiling screen libraries. B) Mean percentage GC content of tiling screen libraries. C) gRNA count in log10 of tiling screen libraries. D) Percentage mapped and unmapped reads of tiling screen libraries. E) Missed gRNA read counts in log10 of tiling screen libraries. F) Gini index of tiling screen libraries. G-I) PCA plots of principal components 1-3 of tiling screen libraries, principal components indicated per plot. J) Sample-to-sample distance plot of tiling screen libraries.

### 7.9.2.2 Expression of genes covered in *Tfap2c* tiling screen



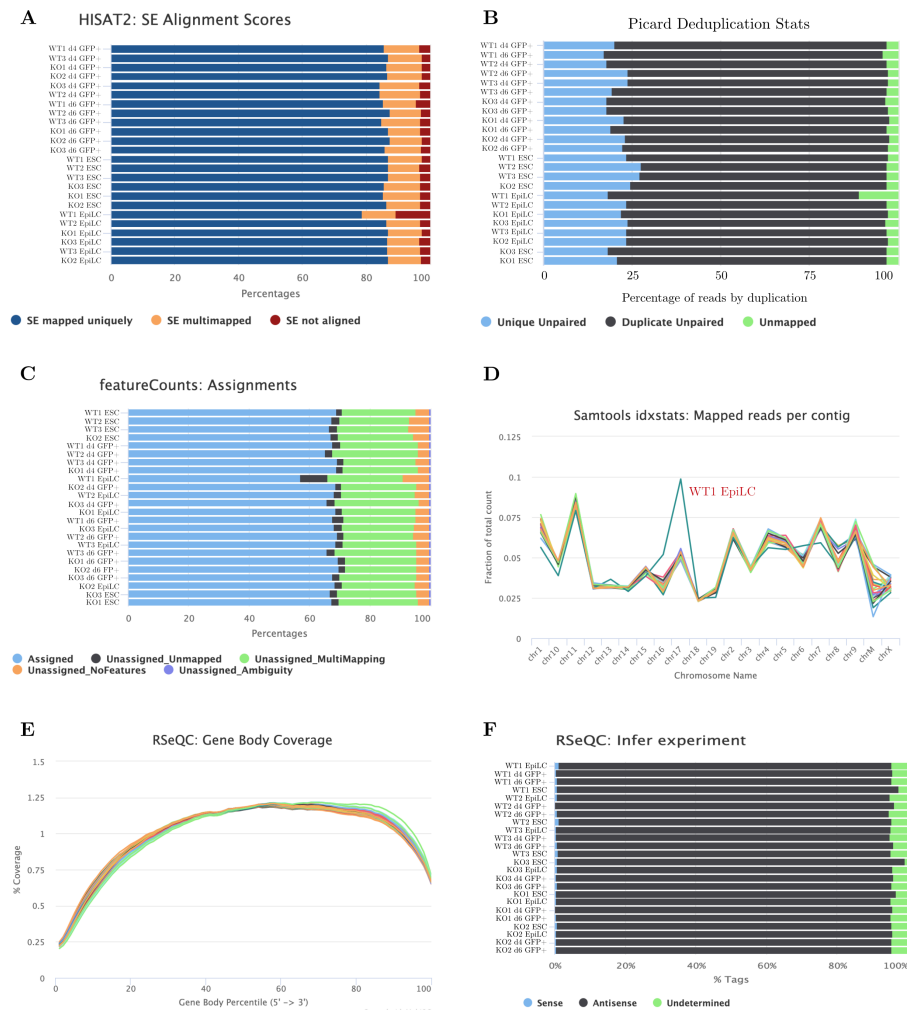
1 = Differentially expressed between WT EpiLCs and WT ESCs as determined by DESeq2 (adj.  $P < 0.05$ )  
 2 = Differentially expressed between WT d4 PGCLCs and WT EpiLCs as determined by DESeq2 (adj.  $P < 0.05$ )

**Figure 56: Heatmaps of genes tiled with gRNAs in *Tfap2c* tiling screen.** Heatmap indicates expression of genes tiled in *Tfap2c* tiling screen over WT differentiation from ESC towards PGCLC fate. 1 indicates significant differential expression (adj.  $p. < 0.05$ ) between ESC and EpiLCs, 2 indicates significant differential expression between EpiLCs and d4 PGCLCs. Red indicates upregulation, blue indicates downregulation.

*Tfap2c* tiling screen library statistics and sample similarity information were generated in MaGECK-VISPR following MaGECK MLE analysis.

## 7.9.3 RNA-sequencing

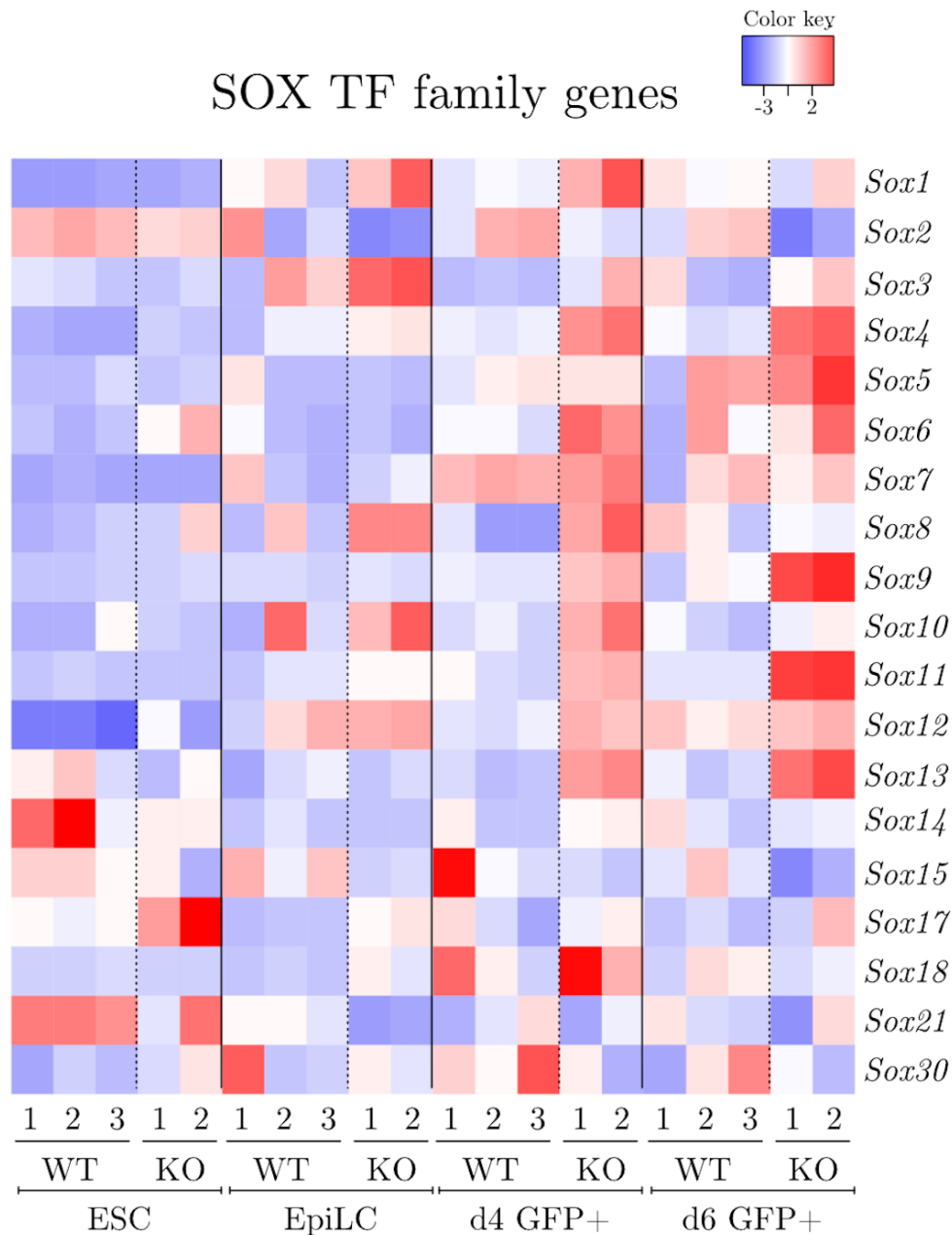
### 7.9.3.1 Direct comparisons of RNA-sequencing libraries



**Figure 57: Quality controls of RNA-seq libraries.** All libraries were aggregated per test with MultiQC. A) HISAT2 genomic alignment of libraries, with color indicating alignment status. Key included. B) Picard deduplication statistics of libraries, where color indicates duplication status. Key included. C) featureCounts assignments of libraries, where color indicates type of feature. Key included. D) SAMtools idxstats demonstrating fraction of total reads mapped per chromosome. Color indicates library, with outlier library WT1 EpiLC indicated in red text. E) RSeQC gene body coverage plot, demonstrating coverage of reads over gene body from 5' to 3' end, color indicates library. F) RSeQC infer experiment analysis of libraries, where color indicates sense of mapped strand. Key included.

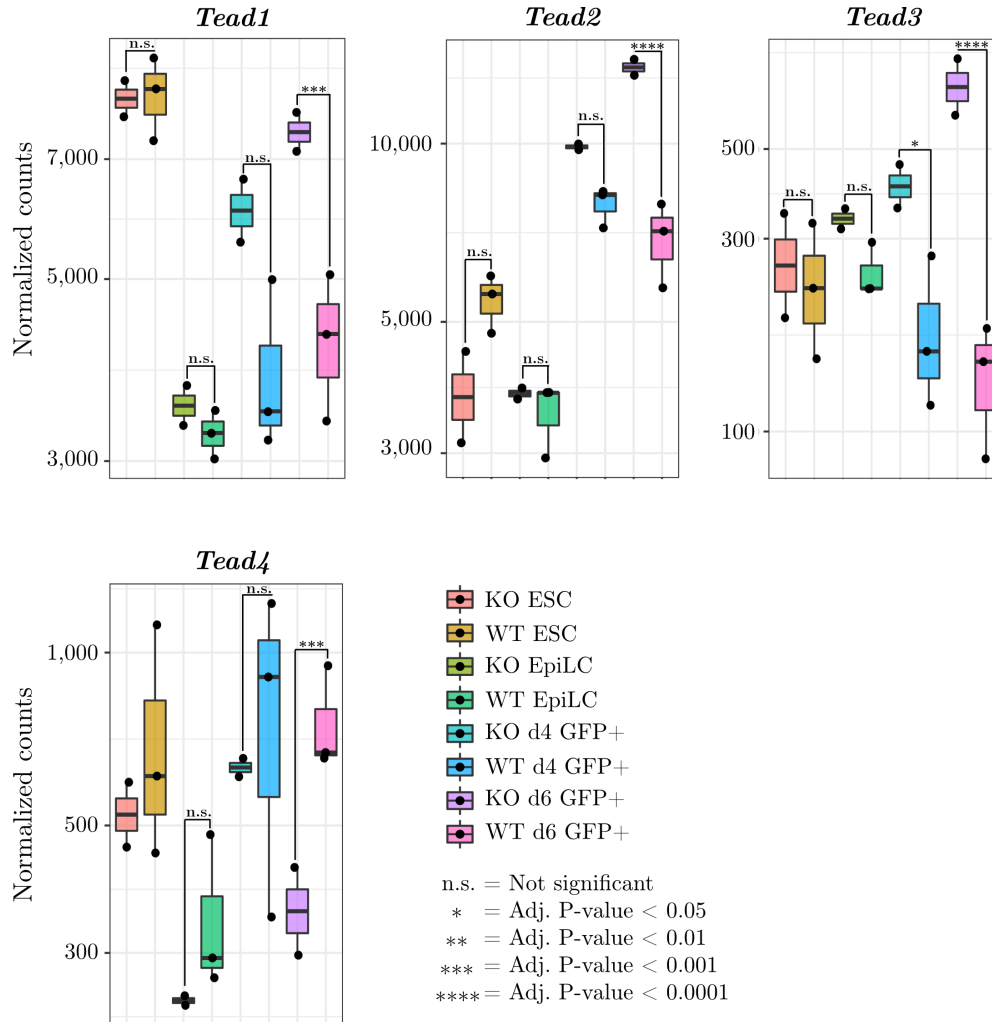
Qualitative assessment of RNA-seq libraries was performed by testing the libraries for genomic alignment, deduplication status, feature assignment, chromosome-mapping, gene-body coverage analysis, and strand-  
edness inference. All libraries demonstrated good gene body coverage and were correctly indicated as  
antisense. All libraries except WT1 EpiLC demonstrated unique genomic mapping >80%, >60% of fea-  
tures were uniquely assigned, and showed close chromosomal fractional counts. WT1 EpiLC demonstrated  
somewhat lower statistics for genomic mapping and deduplication, and also demonstrated a spike in chro-  
mosome 17 reads. All samples demonstrated high levels of duplication. Because all samples in PCA and  
Euclidian sample distance-to-distance clustered correctly (Figure 11), with the exception of KO3 samples  
which were excluded for clustering with WT, sample WT1 EpiLC was not excluded.

### 7.9.3.2 Expression of SOX TF family genes with development following KO of AP2 $\gamma$



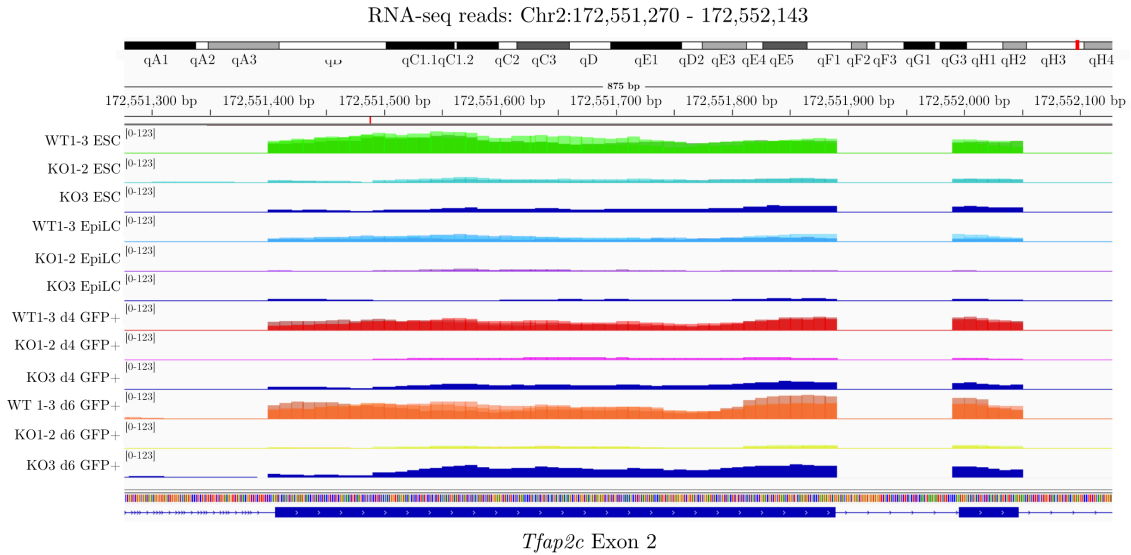
**Figure 58: Expression of SOX TF family genes increases with development towards PGCLC fate following KO of AP2 $\gamma$ .** Heatmap of SOX TF family genes. Red indicates increased expression, blue indicates downregulation. Significance was not determined. Sample and gene names are indicated.

### 7.9.3.3 Expression of TEAD TF family genes with development following KO of AP2 $\gamma$



**Figure 59: TEAD markers are differentially expressed in d6 GFP+ KO cells.** Expression of each gene is indicated by normalized counts (log<sub>10</sub> scale). Significance was determined between AP2 $\gamma$ -KO and WT samples of a given cell type using DESeq2. The color of each bar indicates the sample type, and is written in the key on the bottom right. N = 3 for WT samples, N = 2 for KO samples.

### 7.9.3.4 IGV visualization of cell line KO3



**Figure 60: Visualization of mapped RNA-seq reads on *Tfap2c* exon 2.** Bigwig files generated from RNA-seq mapped reads are visually represented using IGV. WT1-3 and KO1-2 reads are overlaid by cell type, and are indicated to the left of each channel. The site at which KO3 resumes transcription of *Tfap2c* is indicated by the red bar above the reads and adjacent to the DNA distance scale bar. Chromosome 2 is indicated at the top of the scheme, and the red bar indicates the location of *Tfap2c* exon 2 within the chromosome.

To explore why the KO3 d4/d6 GFP+ samples clustered with the WT samples in the PCA analysis, the mapped RNA-seq reads were converted to Bigwig files and visualized using Interactive Genome Viewer (IGV). Examination of *Tfap2c* exon 2 indicates that the number of mapped reads is dramatically reduced in KO1 and KO2 samples relative to WT in every cell type (Appendix 7.9.3.4). Mapped reads appear to be depleted in all KO3 cell types at the 5' end of exon 2, however reads re-appear within the exon and appear to result in reduced transcription relative to the WT but substantially higher than the other KO lines. The site of transcription re-initiation is indicated with a red marking adjacent to the scale bar, and appears to coincide with the site of the 1 bp mutation seen on one allele of KO3 *Tfap2c* exon 2 (Figure 8). As *Tfap2c*-specific reads were clearly mapped in the KO3 samples, and the d4/d6 GFP+ KO3 samples clustered with the WT d4/d6 GFP+ samples, all KO3 samples were excluded from further analysis.

## 7.10 Synthetic DNA sequences

### 7.10.1 Oligos

#### 7.10.1.1 CRISPR/Cas9 knockout oligos

**Table 25:** CRISPR/Cas9 knockout gRNAs designed during this work

<b>gRNA name:</b>	<b>Sequence:</b>	<b>Source:</b>
<i>Tfap2c</i> exon 2-targeting gRNA #2 forward	CACCGCATTTTCTCCATCGTCTT-GCTGG	[165]
<i>Tfap2c</i> exon 2-targeting gRNA #2 reverse	AAACCCAGCAAGACGATGGAGAA-AATGC	[165]
<i>Tfap2c</i> exon 2-targeting gRNA #3 forward	CACCGAAGTAAAATGGCTGCAA-GGAGG	[165]
<i>Tfap2c</i> exon 2-targeting gRNA #3 Reverse	AAACCCTCCTTGCAGCCATTTT-ACTTC	[165]
<i>Tfap2c</i> STOP codon-targeting gRNA for insertion of pm <i>Tfap2c</i> repair #5 forward	CACCGTGGAGAAAATGGAAA-AGCAC	[182]
<i>Tfap2c</i> STOP codon-targeting gRNA for insertion of pm <i>Tfap2c</i> repair #5 reverse	AAACGTGCTTTTCATTTTC-TCCAC	[182]



### 7.10.1.2 pmTfap2c Gibson assembly oligos

**Table 26:** pmTfap2c Gibson assembly oligos designed during this work

Oligo name:	Sequence:
<i>Tfap2c</i> 5' homology Gibson forward	TCAGTGCACAAAGCTTTTATTTCGGTCAAGAGCTGA
<i>Tfap2c</i> 5' homology Gibson reverse	AGCAGACTTCCTCTGCCCTCCTTCCTGTG- CTTTTCCATTT
<i>Tfap2c</i> 5' homology into tdTomato Gibson forward	AAATGGAAAAGCACAGGAAGGAGGGCAGAG- GAAGTCTGCT
tdTomato into selection homology Gibson reverse	TCTTGCAGCTCGGTGACCCG
Selection cassette into <i>Tfap2c</i> 3' homology Gibson forward	CGGGTCACCGAGCTGCAAGA
<i>Tfap2c</i> 3' into selection cassette homology Gibson reverse	GCCTCCTTGCAGCCATTTTATAACTTTAAA- TAATTGGCAT
<i>Tfap2c</i> 3' homology + STOP Gibson forward	ATGCCAATTATTTAAAGTTATAAAATGGCT- GCAAGGAGGC
<i>Tfap2c</i> 3' homology + STOP Gibson reverse	TCCTCTTGAAAACCACACTGGCTAGAGGGG- CTACATGACC
<i>Tfap2c</i> 3' homology into DTA cassette Gibson forward	GGTCATGTAGCCCCTCTAGCCAGTGTGGTT- TTCAAGAGGA
DTA cassette Gibson reverse	CCCTCTAGACTCGAGCGGCCGCTCTAGAA- CTAGTGGATCC

## 7.10.2 Primers

### 7.10.2.1 Genotyping and plasmid-sequencing primers

All genotyping primers were designed with NCBI primer design [331]. All plasmid-sequence primers were designed by Cera McDonald. The U6 sequencing primer was provided by Dr. Ufuk Günesdogan.

**Table 27:** Genotyping primers designed during this work

Primer No.:	Primer name:	Sequence:	Purpose:
G1	pMiniT 2.0 forward	ACCTGCCAACCAAAGCGAGAAC	pMiniT 2.0 validation sequencing
G2	pMiniT 2.0 reverse	TCAGGGTTATTGTCTCATGAGCG	pMiniT 2.0 validation sequencing
G3	pmTfap2c <i>Tfap2c</i> 5' seq reverse 1	TGTATGTTCCGGCTCCAAGAC	pmTfap2c validation sequencing
G4	pmTfap2c <i>Tfap2c</i> 5' seq forward 1	CATCAAGATCGGACACCCAA	pmTfap2c validation sequencing
G5	pmTfap2c <i>Tfap2c</i> 5' seq reverse 2	CTTCCTGTGCTTTTCCATTT	pmTfap2c validation sequencing
G6	pmTfap2c tdTomato seq reverse 1	GCCCTTGCTCACCATTGGGC	pmTfap2c validation sequencing
G7	pmTfap2c T2A seq forward 1	GAGGAAGTCTGCTAACATGC	pmTfap2c validation sequencing
G8	pmTfap2c tdTomato seq reverse 2	TCTTTGATGACGGCCATGT	pmTfap2c validation sequencing
G9	pmTfap2c tdTomato seq forward 2	ACATGGCCGTCATCAAAGA	pmTfap2c validation sequencing
G10	pmTfap2c tdTomato seq reverse 3	CCATGCCGTACAGGAACAG	pmTfap2c validation sequencing

<b>Primer No.:</b>	<b>Primer name:</b>	<b>Sequence:</b>	<b>Purpose:</b>
G11	pmTfap2c tdTomato seq forward 3	CTGTTCCCTGTACGGCATGGA	pmTfap2c validation sequencing
G12	pmTfap2c PGK promoter seq reverse 1	AAAGGCCCGGAGATGAGGAA	pmTfap2c validation sequencing
G13	pmTfap2c PGK promoter seq forward 1	GCTTCAAAAAGCGCACGTCTG	pmTfap2c validation sequencing
G14	pmTfap2c dTK seq reverse 1	CCAGTAAGTCATCGGCTCGG	pmTfap2c validation sequencing
G15	pmTfap2c dTK seq forward 1	ACGATATCGTCTACGTACCC	pmTfap2c validation sequencing
G16	pmTfap2c dTK seq reverse 2	CGATTGGTCGTAATCCAGGA	pmTfap2c validation sequencing
G17	pmTfap2c dTK seq forward 2	CCGTCCCATGCACGTCTTTA	pmTfap2c validation sequencing
G18	pmTfap2c <i>Tfap2c</i> 3' seq reverse 1	CCTTTCAAGAGAGGGAACGG	pmTfap2c validation sequencing
G19	pmTfap2c <i>Tfap2c</i> 3' seq forward 1	CCGTTCCCTCTCTTGAAAGG	pmTfap2c validation sequencing
G20	pmTfap2c <i>Tfap2c</i> 3' seq reverse 2	GCTAGAGGGGCTACATGACC	pmTfap2c validation sequencing
G21	pmTfap2c <i>Tfap2c</i> 3' seq forward 2	GGTCATGTAGCCCCCTTAGC	pmTfap2c validation sequencing
G22	pmTfap2c DTA seq forward 1	CGTGGCCAAGATGCGATGTA	pmTfap2c validation sequencing
G23	pmTfap2c DTA seq reverse 1	CGCCTGACACGATTTCTGTC	pmTfap2c validation sequencing
G24	pmTfap2c DTA seq reverse 2	cggccGCTCTAGAAGTAg	pmTfap2c validation sequencing

<b>Primer No.:</b>	<b>Primer name:</b>	<b>Sequence:</b>	<b>Purpose:</b>
G25	pX330 reverse	CGGGTACCTCTAGAGCCATTT	pX330 validation sequencing
G26	Rox forward 1	ATGCCAATTATTTAAAGTTA	Rox site insertion genotyping
G27	T2A/ <i>Tfap2c</i> exon 7 knock-in reverse	AGCAGACTTCCTCTGCCCTC	To determine if T2A sequence was inserted downstream of <i>Tfap2c</i> exon 7
G28	<i>Tdtomato</i> forward 1	ATGGTGAGCAAGGGCGAGGA	tdTomato gene insertion genotyping
G29	<i>Tdtomato</i> reverse 1	CTACTTGACAGCTCGTCCA	tdTomato gene insertion genotyping
G30	<i>Tdtomato</i> reverse 2	CATGAACTCTTTGATGACCT	tdTomato gene insertion genotyping
G31	<i>Tfap2c</i> exon 7 forward 1	TTTATTCGGTCAAGAGCTGA	Binds in <i>Tfap2c</i> exon 7, to be used to genotype WT or KI T2A-tdTomato insertion
G32	<i>Tfap2c</i> exon 7 reverse 1	TGTATGTTCCGGCTCCAAGAC	Binds in <i>Tfap2c</i> exon 7, to be used to genotype WT allele
G33	<i>Tfap2c</i> 3' UTR reverse 1	GCTAGAGGGGCTACATGACC	Binds in <i>Tfap2c</i> 3' UTR, to be used to genotype WT or KI T2A-tdTomato insertion
G34	<i>Tfap2c</i> 3' UTR reverse 2	CCTTTCAAGAGAGGGAACGG	Binds in <i>Tfap2c</i> 3' UTR, to be used to genotype WT or KI T2A-tdTomato insertion
G35	TRE3G forward 1	gtcttcaagaattcctcgag	pPB-TRE3G m <i>Tfap2c</i> isoform-independent validation sequencing
G36	TRE3G forward 2	agagctcgtttagtgaacctg	pPB-TRE3G m <i>Tfap2c</i> isoform-independent validation sequencing

<b>Primer No.:</b>	<b>Primer name:</b>	<b>Sequence:</b>	<b>Purpose:</b>
G37	TRE3G forward 3	taagcttcgccacatggga	pPB-TRE3G mTfap2c isoform-independent validation sequencing
G38	TRE3G reverse	tcccatggtggcgaagctta	pPB-TRE3G mTfap2c isoform-independent validation sequencing
G39	U6 sequencing primer	GTACAAAATACGTGACGTAG	Sequencing insertion of gRNA into pX330 or pKLV2-W plasmids

### 7.10.2.2 NGS sequencing primers

All ATAC primers are custom Nextera sequencing primers [332]. All RNA-seq sequencing primers are from the NEBNext® multiplex oligos for Illumina® index primer sets 2-3 (New England BioLabs). *Tfap2c* tiling screen (TTS) primers were taken from Hackett *et al.*, 2018 [172]. All indexing sequences are indicated in bold. In the following NEBNext primer sequences, -s-T indicates a phosphorothioate bond. 5phos indicates a 5' phosphorylation modification.

**Table 28:** Sequencing primers used during this work

Primer name:	Sequence:
Nextera Ad1	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCA GGCTCAGATGTG
Nextera Ad2.1	CAAGCAGAAGACGGCATAACGAGAT <b>TGCGCT</b> TAGTCTCGTG GGCTCGGAGATGT
Nextera Ad2.2	CAAGCAGAAGACGGCATAACGAGAT <b>CTAGTACGGT</b> TCTCGTG GGCTCG GAGATGT
Nextera Ad2.3	CAAGCAGAAGACGGCATAACGAGAT <b>TTCTGCT</b> TGTCTCGTG GGCTCGGAGATGT
Nextera Ad2.4	CAAGCAGAAGACGGCATAACGAGAT <b>GCTCAGGAGT</b> TCTCGTG GGCTCGGAGATGT
Nextera Ad2.5	CAAGCAGAAGACGGCATAACGAGAT <b>AGGAGTCCGT</b> TCTCGTG GGCTCGGAGATGT
Nextera Ad2.6	CAAGCAGAAGACGGCATAACGAGAT <b>CATGCC</b> TAGTCTCGTG GGCTCGGAGATGT
Nextera Ad2.7	CAAGCAGAAGACGGCATAACGAGAT <b>GTAGAGAGG</b> TCTCGTG GGCTCGGAGATGT
Nextera Ad2.8	CAAGCAGAAGACGGCATAACGAGAT <b>CCTCTCTG</b> GTCTCGTG GGCTCGGAGATGT
Nextera Ad2.9	CAAGCAGAAGACGGCATAACGAGAT <b>AGCGTAGCGT</b> TCTCGTG GGCTCGGAGATGT
Nextera Ad2.10	CAAGCAGAAGACGGCATAACGAGAT <b>CAGCCTCGG</b> TCTCGTG GGCTCGGAGATGT
Nextera Ad2.11	CAAGCAGAAGACGGCATAACGAGAT <b>TGCTCTT</b> TGTCTCGTG GGCTCG GAGATGT
Nextera Ad2.12	CAAGCAGAAGACGGCATAACGAGAT <b>TCCTCTAC</b> GTCTCGTG GGCTCGGAGATGT
Nextera Ad2.13	CAAGCAGAAGACGGCATAACGAGAT <b>ATCACGAC</b> GTCTCGTG GGCTCGGAGATGT
Nextera Ad2.14	CAAGCAGAAGACGGCATAACGAGAT <b>ACAGTGGT</b> GTCTCGTG GGCTCGGAGATGT
Nextera Ad2.15	CAAGCAGAAGACGGCATAACGAGAT <b>CAGATCCAGT</b> TCTCGTG GGCTCGGAGATGT

<b>Primer name:</b>	<b>Sequence:</b>
Nextera Ad2.16	CAAGCAGAAGACGGCATAACGAGAT <b>ACAAACGGGTCTCGTG</b> GGCTCGGAGATGT
Nextera Ad2.17	CAAGCAGAAGACGGCATAACGAGAT <b>ACCCAGCAGTCTCGTG</b> GGCTCGGAGATGT
Nextera Ad2.18	CAAGCAGAAGACGGCATAACGAGAT <b>AACCCCTCGTCTCGTG</b> GGCTCGGAGATGT
Nextera Ad2.19	CAAGCAGAAGACGGCATAACGAGAT <b>CCCAACCTGTCTCGTG</b> GGCTCGGAGATGT
Nextera Ad2.20	CAAGCAGAAGACGGCATAACGAGAT <b>CACCACACGTCTCGTG</b> GGCTCGGAGATGT
Nextera Ad2.21	CAAGCAGAAGACGGCATAACGAGAT <b>GAAACCCAGTCTCGTG</b> GGCTCGGAGATGT
Nextera Ad2.22	CAAGCAGAAGACGGCATAACGAGAT <b>TGTGACCAGTCTCGTG</b> GGCTCGGAGATGT
Nextera Ad2.23	CAAGCAGAAGACGGCATAACGAGAT <b>AGGGTCAAGTCTCGTG</b> GGCTCGGAGATGT
Nextera Ad2.24	CAAGCAGAAGACGGCATAACGAGAT <b>AGGAGTGGGTCTCGTG</b> GGCTCGGAGATGT
NEBNext Adapter for Illumina	5Phos-GATCGGAAGAGCACACGTCTGAACTCCAGTCUACAC TCTTTCCCTACACGACGCTCTTCCGATC-s-T
NEBNext Index 16 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATG <b>CGGACGGTGACTGGAGTT</b> CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 17 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGAT <b>CTCTACGTGACTGGAGTT</b> CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 18 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATG <b>TGGACGTGACTGGAGTT</b> CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 19 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATCG <b>TTTACGTGACTGGAGTT</b> CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 21 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATT <b>CCGAAACGTGACTGGAGTT</b> CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 22 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATT <b>ACGTACGGTGACTGGAGTT</b> CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 23 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATAT <b>CCACTCGTGACTGGAGTT</b> CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 24 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATG <b>CTACCGTGACTGGAGTT</b> CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 25 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATAT <b>ATCAGTGTGACTGGAGTT</b> CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 26 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATG <b>CTCATGTGACTGGAGTT</b> CAGACGTGTGCTCTTCCGATC-s-T

<b>Primer name:</b>	<b>Sequence:</b>
NEBNext Index 27 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATAA <b>AGGAAT</b> GTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 28 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATC <b>TTTTG</b> GTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 29 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGAT <b>TAGTTG</b> GTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 31 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGAT <b>ATCGT</b> GTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 32 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGAT <b>TGAGT</b> GTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 33 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGAT <b>CGCCT</b> GTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T-3
NEBNext Index 34 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGAT <b>GCCAT</b> GTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 35 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGATA <b>AAAAT</b> GTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 36 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGAT <b>TGTTG</b> GTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 42 Primer for Illumina	CAAGCAGAAGACGGCATAACGAGAT <b>CGATT</b> GTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T
NEBNext Universal PCR Primer for Illumina	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATC-s-T
TTS 79mer L1	GCGACGAGAAGACTAAAAC
TTS 79mer U1	GCAGATGGCTCTTTGTCCTA
TTS Universal forward	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTCTTGTGGAAAGGACGAAACAC
TTS Universal forward +1 stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCTTCTTGTGGAAAGGACGAAACAC
TTS Universal forward + 2 staggers	AATGATACGGCGACCACCGAGATCTACACTCTTTCC CTACACGACGCTCTTCCGATCTATCTTGTGGAAAGGACGAAACAC
TTS Index 2 reverse	CAAGCAGAAACGGCATAACGAGATACATCGGTGACTG GAGTTCA- GACGTGTGCTCTTCCGATCCTAAAGCGCATGCTCCAGAC
TTS Index 5 reverse	CAAGCAGAAGACGGCATAACGAGATCACTGTGTGAC TGGAGTTCA- GACGTGTGCTCTTCCGATCTAAAGCGCATGCTCCAGAC
TTS Index 6 reverse	CAAGCAGAAGACGGCATAACGAGATATTGGCGTGA CTGGAGTTCA- GACGTGTGCTCTTCCGATCTAAAGCGCATGCTCCAGAC
TTS Index 12 reverse	CAAGCAGAAGACGGCATAACGAGATFACAAGGTG ACTGGAGTTCA- GACGTGTGCTCTTCCGATCTAAAGCGCATGCTCCAGAC
TTS Index 19 reverse	CAAGCAGAAGACGGCATAACGAGATCGTTTCAC GTGACTG- GAGTTCAGACGTGTGCTCTTCCGATCTAAAGCGCATGCTCCAGAC



### 7.10.2.3 qPCR primers

Primers indicated with an asterisk (\*) following the name were provided by Dr. Ufuk Günesdogan, otherwise all primers were designed using NCBI primer design [331].

**Table 29:** qPCR primers used and designed during this work

<b>Primer name:</b>	<b>Sequence:</b>
Arbp forward*	CAAAGCTGAAGCAAAGGAAGAG
Arbp reverse*	AATTAAGCAGGCTGACTTGGTTG
Dazl forward*	CCAGAAGGCAAAATCATGCCAA
Dazl reverse*	GGCAAAGAAACTCCTGATTTCCG
Dnmt3b forward*	CTCGCAAGGTGTGGGCTTTTGTAAC
Dnmt3b reverse*	CTGGGCATCTGTCATCTTTGCACC
Dppa3 forward*	TCCGATTGAGCAGAGACAAA
Dppa3 reverse*	GATTTCCCAGCACCAGAAAA
Prdm1 forward	CCGCGGCCGTAGAAAAGGA
Prdm1 reverse	TTGGGGGCAGCCAAGGTC
Prdm14 forward*	GCCTGAACAAGCACATGAGA
Prdm14 reverse*	TGCACTTGAAGGGCTTCTCT
Sox2 forward	GGAAAGGGTTCTTGCTGGGT
Sox2 reverse	ACGAAAACGGTCTTGCCAGT
Tfap2a forward	CCAACGTTACCCTCCTCACG
Tfap2a reverse	TTCGCACACGTACCCAAAGT
Tfap2c forward*	CGCGGAAGAGTATGTTGTTG
Tfap2c reverse*	CGATCTTGATGGAGAAGGTCA

### 7.10.3 gBlocks

Three gBlocks gDNA fragments were ordered from IDT to facilitate cloning of *Tfap2c* isoforms 1 and 2 cDNA into the pPB-TRE3G-IRES-mC2 plasmid (see Section ??).

The sequence of the unique *Tfap2c* isoform 1 exon 1 and overlap with the pPB-TRE3G-IRES-mC2 plasmid and the 5' *Tfap2c* conserved mRNA sequence is as follows:

```
ccacttctaccctcgtaaagtcgacgacctccatagaagacaccgggaccgatccagcctccggactctagcggtttaaacttaagcttcgccacc
atgggatccATGTTGTGGAAAATAACAGATAATGTCAAGTATGAAGAGGATTGCGAGGATCGCCACGACTCGAGC
```

The sequence in red represents the 5' overhang sequence homologous to the plasmid. The sequence in purple represents the overhang sequence homologous to the *Tfap2c* conserved mRNA sequence.

The sequence of the unique *Tfap2c* isoform 2 exon 1 and overlap with the pPB-TRE3G-IRES-mC2 plasmid and the 5' *Tfap2c* conserved mRNA sequence is as follows:

```
ccacttctaccctcgtaaagtcgacgacctccatagaagacaccgggaccgatccagcctccggactctagcggtttaaacttaagcttcgccacc
atgggatccATGGGAGGAGGGTTGTTGAACGAGGGCGCTGTGCGCCGAGGTCGCGCACGCCGAACCCCTGGGGCACGGTAACCGACTGTGGCG
CGGAACGCGCTCTTGGCACCTCCAAGCGACGCAGCGCGGAACCTTCGCCAGAGGGCTCGAAGGATTGTCGCCCTCAGGATCGCCACGACTCGAGC
```

The sequence of the conserved *Tfap2c* cDNA sequence encoding exons 2-7 shared by both isoforms and the overlap with the pPB-TRE3G-IRES-mC2 plasmid is as follows:

```
AGGATCGCCACGACTCGAGCAGTAATGGCAACCCTCGCATCCCTCACCTCTCCTCTCCGGACAACATCTCTACAGTCCCAGCCGCGCTCTCTCGC
ACACCGGGGTTGCAGAGTACCAGCCGCTCCTTACTTCCCAGCCGCTTACCAGCAGTGGCATACTCGAGTCCGCGACCACTTACTCGCATCTGG
GAGAGGCTTACGCTGCCCATGAACCCCTGCACCAGCCTGCGGCCACCGGCAGCCAGCAACAGGCCTGGCCGGGTCGACAGAGTCAGAGGGCT
CTAGCTGGCTCGACCACAGCCGCTCTGCAAGTCTAATACCCATATTTTCAGGGCTGGAGGGGGCTCGGTGAGCGCCCGCGGGAAGTCTACC
GCCGTCGACCTGCTGCTGCCTCACGCGCAGCCCTGGAAGCCGCGCTGGCTGAGAACCTGGGGCTGCACGAGATGGCTCACCCATAGAGGAGG
TGCAGAATGTGGACGACGCGCACTTGTCTTACACGATCAGACTGTCATTTCGCAAGGACCCATTTTCGATGACCAAGAACCCTTTGGGGCTCCCTT
GCCAGAAGGACCTGGTGGGAGTGGTCATGAACCCAGTGGTCTTCTGCTCGTCCCTGGAAGACTGTCCCTGCTCAGTCCACGTGCAAGTACA
AAGTAAGTGTGGCTGAGGTACAGAGGCGACTGTACCACCGGAATGCCTAAACGCTCGTCTCTGGGAGGTGTGCTCAGAAGAGCAAAGTCCAAAA
ATGGAGGCCGCTCTTGAGGGAGAAGTTGGACAAAATTGGATTGAACCTTCCGCGCGGAGACGAAAGCTGCCACGTCACCTCTCTCACGTCTC
TCGTGGAAGGTGAAGCCGTCCACCTAGCACGGACTTCGCCTATGTCTGCAAGCTGAGTCCCTAGTAAAGCGGTGGCTGACTATTTAACGAGAC
CACATCTTGGGGACGGAATGAGATGGCCACGCGGAAGAGTATGTTGTTGGCTGCACAGCAGGTGTGCAAGGAGTTCACTGACCTTCTCCATCAAG
ATCGGACACCAACGGGAACAACAGGCCCGCCAGGCTTTGGAGCCGAACATACAAAAGTGTGTTGCTCATTTCAGCCTGATAACTCATGGCTTTG
GCAGCCAGGCCATCTGTGCGCGGTCTCCGAGTGCAGAATTATATCAAGGAGGCTTAATCGCCATCGATAAGTCTACATGAACCCGGGGACC
AGAGTCCGGCTGATTCCAGCAAGACGATGGAGAAAATGGAAAAGCACAGGAAGTAAGCTAGCGGCGCTCGATAAGC
```

The sequence in purple indicates overlap with the unique isoform 1 and isoform 2 gBlock sequences, while the sequence in green indicates the 3' overhang sequence homologous with the pPB-TRE3G-IRES-mC2 plasmid.

## 7.11 *Tfap2c* tiling screen gRNA design

**Table 30:** Genomic regions and genes targeted during the *Tfap2c* tiling screen

Targeted region:	Genomic coordinates:	Designer of gRNAs:	Designed using:	Reason for targeting:
TAD 325	chr2:172,454,500-172,814,499	Xiaojuan Li	Joung <i>et al.</i> script	TAD of <i>Tfap2c</i>
Partial TAD 324	chr2:171,904,500-172,454,500	Xiaojuan Li	Joung <i>et al.</i> script	Contains final quarter of TAD 324 (upstream of TAD 325), as well as continuous sequence to TAD 325; acts as negative control
Partial TAD 326	chr2:172,814,499-173,624,499	Xiaojuan Li	Joung <i>et al.</i> script	Contains first quarter of TAD 326 (downstream of TAD 325), as well as continuous sequence from TAD 325; acts as negative control
Potential regulatory region 1	chr2:170,246,202-170,503,872	Xiaojuan Li	Joung <i>et al.</i> script	Two putative <i>Tfap2c</i> -interacting CREs were identified in this region in unpublished Hi-C data
Potential regulatory region 2	chr2:158,575,739-158,597,087	Xiaojuan Li	Joung <i>et al.</i> script	One putative <i>Tfap2c</i> -interacting CRE was identified in this region in unpublished Hi-C data
Potential regulatory region 3	chr2:178,627,369-178,633,710	Xiaojuan Li	Joung <i>et al.</i> script	One putative <i>Tfap2c</i> -interacting CRE was identified in this region in unpublished Hi-C data
Potential regulatory region 4	chr11:78,007,245-78,013,145	Xiaojuan Li	Joung <i>et al.</i> script	One putative <i>Tfap2c</i> -interacting CRE was identified in this region in unpublished Hi-C data
Potential regulatory region 5	chr1:100,238,852-100,242,467	Xiaojuan Li	Joung <i>et al.</i> script	One putative <i>Tfap2c</i> -interacting CRE was identified in this region in unpublished Hi-C data

<b>Targeted region:</b>	<b>Genomic coordinates:</b>	<b>Designer of gRNAs:</b>	<b>Designed using:</b>	<b>Reason for targeting:</b>
Positive control enhancer 1	chr5:119,656,454-119,661,324	Xiaojuan Li	Joung <i>et al.</i> script	Validated ESC enhancer included as positive control
Positive control enhancer 2	chr8:43,351,590-43,356,510	Xiaojuan Li	Joung <i>et al.</i> script	Validated ESC enhancer included as positive control
Positive control enhancer 3	chr14:99,323,583-99,327,543	Xiaojuan Li	Joung <i>et al.</i> script	Validated ESC enhancer included as positive control
Positive control enhancer 4	chr8:72,290,869-72,293,289	Xiaojuan Li	Joung <i>et al.</i> script	Validated ESC enhancer included as positive control
Negative control enhancer 1	chr8:43,327,503-43,332,433	Cera McDonald	Benchling	Invalidated ESC enhancer included as negative control
Negative control enhancer 2	chr14:99,333,716-99,341,806	Cera McDonald	Benchling	Invalidated ESC enhancer included as negative control
Negative control enhancer 3	chr17:47,950,985-47,959,425	Cera McDonald	Benchling	Invalidated ESC enhancer included as negative control
Negative control enhancer 4	chr18:81,196,705-81,198,815	Cera McDonald	Benchling	Invalidated ESC enhancer included as negative control
<i>Dppa3</i> gene	chr6:122,626,410-122,630,272	Joung <i>et al.</i>	Joung <i>et al.</i> script	Positive control targeting <i>Dppa3</i> , a PGC marker gene as well as the gene reported by eGFP expression in the SGET cell line
<i>Dppa5a</i> gene	chr9:78,367,052-78,368,177	Joung <i>et al.</i>	Joung <i>et al.</i> script	Negative control targeting <i>Dppa5a</i> gene, which should not be enriched in PGCs and is reported by tdTomato expression in the SGET cell line
<i>eGFP</i> gene	N/A	Shalem <i>et al.</i>	Shalem <i>et al.</i> , 2014 script	Positive control targeting fluorescent reporter of PGC gene <i>Dppa3</i>
<i>Prdm1</i> gene	chr10:44,438,293-44,458,687	Joung <i>et al.</i>	Joung <i>et al.</i> script	Positive control targeting key PGC gene <i>Prdm1</i>

<b>Targeted region:</b>	<b>Genomic coordinates:</b>	<b>Designer of gRNAs:</b>	<b>Designed using:</b>	<b>Reason for targeting:</b>
<i>Prdm14</i> gene	chr1:13,113,457-13,127,163	Joung <i>et al.</i>	Joung <i>et al.</i> script	Positive control targeting key PGC gene <i>Prdm14</i>
<i>tdTomato</i> gene	N/A	Cera McDonald	Benchling	Negative control targeting reporter of non-PGC gene <i>Dppa5a</i>
<i>Tfap2c</i> gene	chr2:172,549,593-172,558,622	Joung <i>et al.</i>	Joung <i>et al.</i> script	Positive control targeting key PGC gene <i>Tfap2c</i>
Negative control region 1	chr9:29,055,657-29,060,656	Cera McDonald	Benchling	Negative control region with no known relationship to expression or regulation of <i>Tfap2c</i>
Negative control region 2	chr13:66,700,268-66,705,267	Cera McDonald	Benchling	Negative control region with no known relationship to expression or regulation of <i>Tfap2c</i>
Scrambled gRNAs	N/A	Joung <i>et al.</i>	Joung <i>et al.</i> script	Scrambled gRNA sequences designed not to target mouse genome but to act as negative controls

# List of Figures

- 1 **Early embryonic development in the mouse.** Totipotency is established in the zygote following fertilization, but is lost once asymmetric cell division produces the distinct inner and outer cell populations of the morula by E2.5. Induction of the trophectoderm (TE; indicated in purple) begins in the morula stage with the formation of polarized, compacted cells which begin to express TE factor CDX2. Induction of the pluripotent inner cell mass (ICM) also begins in the morula with the formation of apolar cells which begin to express pluripotency factors OCT4 and NANOG. By E3.5 in the blastocyst, the TE has formed and begins to upregulate TE factor ID2. At this stage the ICM has also formed, and begins to express pluripotency factor SOX2. Modified from Burton & Torres-Padilla, 2014 [4]. . . . . 12
- 2 **Post-implantation development in the early mouse embryo.** The developmental timeline of the mouse embryo is shown from E4.5-7.0. Cells transition from naïve pluripotency at E4.5 through formative pluripotency by E5.0 and primed pluripotency of the posterior-epiblast by E6.5. Cells deriving from the epiblast lineage and which will ultimately give rise to the embryo proper are shown in dark blue. Cells of the TE lineage which will ultimately form or contribute to the placenta are shown in light blue. Cells deriving from the PrE lineage and which give rise to the visceral endoderm (VE) are shown in yellow. The anterior visceral endoderm (AVE) is indicated in pink, while the primitive streak is indicated in purple. The emergence of the primordial germ cells (PGCs) by E7.0 is indicated. Figure taken from Ávila-González *et al.*, 2021 [12]. . . . . 14
- 3 **PGC induction in the mouse.** Mouse PGC precursors (mPGCs; shown in green) are induced following exposure to activating signals BMP4 and 8B (shown in green) from the extraembryonic ectoderm (ExE; shown in beige), BMP2 signals from the proximal visceral endoderm (VE; shown in yellow), and WNT3 signals from the posterior epiblast and posterior VE (both shown in pink). Inhibitory signals including CER1, DKK1, and LEFTY1 released from the anterior VE (AVE; shown in brown) prevent the induction of PGCs, and thus control the location and opportunity for PGC induction. Together, these activating and antagonizing signals ensure a small population of mPGCs may be specified in the proximal-posterior region of the epiblast. The directional axes of the embryo are indicated with P - D (proximal to distal) and A - P (anterior to posterior). The primitive streak (PS) is also indicated. Modified from Tang *et al.*, 2016 [24]. . . . . 16
- 4 **PGCLC induction and comparison to *in vivo* murine tissue.** Induction of PGCLCs begins with the culture of mouse ESCs (mES) in 2i/LIF (2i+LIF) medium. These cells may be induced directly towards mouse EpiLC (mEpiLC) fate for approximately two days in defined medium containing activin A and f FGF2 (Fibroblast growth factor 2; also known as bFGF). The resultant EpiLCs are comparable to cells of the E5.5-5.75 post-implantation epiblast. The EpiLCs may be directly induced towards mouse PGCLC (mPGCLC) fate by aggregating them as EBs in medium containing the cytokines BMP4, BMP8B, SCF, LIF, and EGF. After six days, they will resemble E9.5 migratory PGCLCs both transcriptionally and epigenetically. Modified from Tang *et al.*, 2016 [24]. . . . . 21

5	<b>Predicted protein structure of AP2<math>\gamma</math>.</b> The predicted protein structure of AP2 $\gamma$ isoform 1 generated by AlphaFold Monomer v2.0 [89; 90]. The per-residue confidence score is indicated by color: high confidence - confidence - low - very low is indicated by dark blue - light blue - yellow - orange. . . . .	22
6	<b>Isoforms of AP2<math>\gamma</math> on the DNA and protein level.</b> A) Coding DNA sequence (CDS) of <i>Tfap2c</i> isoform 1 with exons indicated. B) CDS of <i>Tfap2c</i> isoform 2 with exons indicated. C) Alignment of exon 1 and partial exon 2 in CDS of <i>Tfap2c</i> isoforms 1 and 2. D) Amino acid (aa) sequence of AP2 $\gamma$ isoform 1, with features indicated. E) Aa sequence of AP2 $\gamma$ isoform 2, with features indicated. All images were generated using SnapGene <sup>®</sup> software. . . . .	23
7	<b>Histone modifications indicate the activity status of enhancers and promoters.</b> Histone modifications allow for the discrimination of CREs based on activity. Nucleosome-free promoters with flanking regions enriched for H3K27ac and H3K4me3 are classified as active, and nucleosome-free enhancers in with flanking regions enriched for H3K4me1 and H3K27ac are also classified as active. Closed or poised enhancers may be identified by the presence of nucleosomes and the histone modifications H3K27me3 and H3K4me1, while a primed enhancer may be nucleosome-free and flanked with regions enriched for H3K4me1. RNA polymerase II (Pol II), transcription factors (TFs), and DNA binding motifs are indicated. Modified from Shlyueva <i>et al.</i> , 2014 [137]. . . . .	29
8	<b>Validation of B1-based AP2<math>\gamma</math>-KO ESC lines.</b> A) Comparison of translated AP2 $\gamma$ protein sequences between WT B1 ESCs and AP2 $\gamma$ -KO lines KO1, KO2, KO3. Premature stop codons are indicated as asterisks (*), and blue-highlighting demonstrates consensus between the WT and KO amino acid (aa) sequences. B) Western blot of WT, KO1, KO2, and KO3 ESC lysates, with 80 $\mu$ g protein/lane. AP2 $\gamma$ isoforms 1 and 2 are indicated, with expected size in kilodaltons (kDa) of 49 and 53 kDa, respectively. The upper blot indicates $\alpha$ 'AP2 $\gamma$ -HRP labeling, while the lower blot indicates sequential $\alpha$ 'GAPDH-HRP labeling of the same blot, with expected size in kDa of 37 kDa. A protein ladder indicates sizes in kDa in the leftmost lane. C) 40x confocal images depicting the IF staining of WT and KO1, KO2, and KO3 ESCs. $\alpha$ -AP2 $\gamma$ -Alexa-488 is shown in green and DAPI is shown in blue, with scale bar of 100 $\mu$ m indicated. D) 10x phase-contrast light microscopy images of live WT, KO1, KO2, and KO3 ESCs. Inset demonstrates digital zoom of region of interest, and scale bar of 200 $\mu$ m is indicated. . . . .	68
9	<b>The AP2<math>\gamma</math>-KO phenotype is validated in IF stainings of d6 KO1 EBs.</b> WT and KO1 EpiLCs were aggregated as EBs for six days in either -cyto control medium or +cyto medium which induces the cells towards PGCLC fate. The EBs were subsequently fixed, cryosectioned, and labeled with $\alpha$ 'AP2 $\gamma$ -Alexa-555 (shown in red), $\alpha$ 'GFP-Alexa-488 (shown in green), and DAPI. Overlay images of the $\alpha$ 'AP2 $\gamma$ and $\alpha$ 'GFP channels are shown, as are scale bars of 50 $\mu$ m. . . . .	69
10	<b>GFP expression not significantly different between d6 cyto+ WT and AP2<math>\gamma</math>-KO cells.</b> The percentage of GFP+ cells indirectly reporting BLIMP1 expression in B1 WT and KO1, KO2, and KO3 EBs grown in +/- cyto medium is indicated. N = 3 for all samples. Significance of p < 0.05 is indicated, and was determined using the unpaired t-test in Prism 9. . . . .	70

11	<p><b>RNA-seq analysis demonstrates transcriptional differences between cell types as well as WT and AP2<math>\gamma</math>-KO d4/d6 GFP+ cells.</b> A) PCA plots of RNA-seq samples with and without the inclusion of KO3 samples. KO3 d4/d6 GFP+ samples which clustered with WT d4/d6 GFP+ samples are indicated in the red ring, all other cell types and clusters are indicated in black. Color of marker indicates cell type and KO status and is explained in the key, and the variance of principal components (PC1 and PC2) is indicated. B) Scree plot of principal components used to generate PCA plots. The number of principal components and proportion of explained variance is indicated. C) Euclidian sample-to-sample distance of all RNA-seq samples, with genotype (condition) and cell type (type) indicated. Scale indicates relative difference between samples. . . . .</p>	72
12	<p><b>Key PGC markers are differentially expressed in d4/d6 GFP+ KO cells.</b> Expression of each gene is indicated by normalized counts (log10 scale). Significance was determined between AP2<math>\gamma</math>-KO and WT samples of a given cell type using DESeq2, and is indicated in the key. The color of each bar indicates the sample type, and is written in the key on the bottom right. N = 3 for WT samples, N = 2 for KO samples. . . . .</p>	74
13	<p><b>KO of AP2<math>\gamma</math> results in differential developmental gene expression in ESCs.</b> A) Log ratio (M) and mean average (A) plot (MA plot) of AP2<math>\gamma</math>-KO versus WT ESC RNA-seq genes, generated using DESeq2. Differentially expressed (adj. p. &lt; 0.01) genes are indicated in blue. B) Goseq gene ontology results for biological process GO terms overrepresented in DE genes between AP2<math>\gamma</math>-KO and WT ESC samples. Significance is indicated by the color of the marker, and gene counts associated with each cluster are indicated by the size of the marker. C) Heatmap of developmental DE genes. D) Heatmap of imprinted DE genes. All heatmaps were generated with heatmap2, and the general classification of each gene is provided to the right of the gene name, with sample name and KO status indicated below. Red indicates upregulation and blue indicates downregulation. Samples are hierarchically clustered by similarity. . . . .</p>	76
14	<p><b>KO of AP2<math>\gamma</math> results in differential developmental gene expression in EpiLCs.</b> A) MA plot of AP2<math>\gamma</math>-KO EpiLC versus WT RNA-seq genes, generated using DESeq2. Differentially expressed (adj. p. &lt; 0.01) genes are indicated in blue. B) Goseq gene ontology results for biological process GO terms overrepresented in DE genes between AP2<math>\gamma</math>-KO and WT EpiLC samples. Significance is indicated by the color of the marker, and gene counts associated with each cluster are indicated by the size of the marker. C) Heatmap of pluripotency and developmental DE genes. D) Heatmap of imprinted DE genes. E) Heatmap of DE embryonic lineage genes. Heatmaps were generated with heatmap2, and the general classification of each gene is provided to the right of the gene name, with sample name and KO status indicated below. Red indicates upregulation and blue indicates downregulation. Samples are hierarchically clustered by similarity. . . . .</p>	78



15	<b>KO of AP2<math>\gamma</math> impairs the expression of PGC and pluripotency markers in d4 GFP+ cells.</b> A) MA plot of AP2 $\gamma$ -KO d4 GFP+ versus WT RNA-seq genes, generated using DESeq2. Differentially expressed (adj. p. < 0.01) genes are indicated in blue. B) GOseq gene ontology results for biological process GO terms overrepresented in DE genes between AP2 $\gamma$ -KO and WT d4 GFP+ samples. Significance is indicated by the color of the marker, and gene counts associated with each cluster are indicated by the size of the marker. C) Heatmap of pluripotency and developmental DE genes. D) Heatmap of imprinted DE genes. E) Heatmap of DE embryonic lineage genes. Heatmaps were generated with heatmap2, and the general classification of each gene is provided to the right of the gene name, with sample name and KO status indicated below. Red indicates upregulation and blue indicates downregulation. Samples are hierarchically clustered by similarity. . . . .	79
16	<b>KO of AP2<math>\gamma</math> profoundly alters differential expression of developmental genes in d6 GFP+ cells.</b> A) MA plot of AP2 $\gamma$ -KO d6 GFP+ versus WT RNA-seq genes, generated using DESeq2. Differentially expressed (adj. p. < 0.01) genes are indicated in blue. B) GOseq gene ontology results for biological process GO terms overrepresented in DE genes between AP2 $\gamma$ -KO and WT d6 GFP+ samples. Significance is indicated by the color of the marker, and gene counts associated with each cluster are indicated by the size of the marker. . . . .	80
17	<b>Many pluripotency, somatic, and imprinted genes are aberrantly expressed in AP2<math>\gamma</math>-KO d6 GFP+ cells relative to WT.</b> A) Heatmap of PGC and somatic fate DE genes. B) Heatmap of DE pluripotency genes. C) Heatmap of DE embryonic lineage genes. D) Heatmap of DE imprinting genes. Heatmaps were generated with heatmap2, and the general classification of each gene is provided to the right of the gene name, with sample name and KO status indicated below. Red indicates upregulation and blue indicates downregulation. Samples are hierarchically clustered by similarity. . . . .	81
18	<b><i>Tfap2a</i>, <i>Myc</i>, and <i>Cd44</i> are DE in AP2<math>\alpha</math>-KO d6 GFP+ cells.</b> The normalized counts of DE genes of interest <i>Tfap2a</i> , <i>Myc</i> , and <i>Cd44</i> in AP2 $\alpha$ -KO and WT ESCs, EpiLCs, and PGCLCs are shown. Significance between AP2 $\alpha$ -KO and WT samples was determined with DESeq2. WT N = 3, AP2 $\alpha$ -KO N = 2. Normalized counts are shown (log10 scale), and the sample/significance key is included below the plots. . . . .	82
19	<b>AP2<math>\gamma</math>-KO d4/d6 GFP+ cells upregulate many markers of EMP fate.</b> EMP gene expression was examined by determining the significance (adj. p. < 0.01, fc +/- 1.5x) of DE genes in AP2 $\alpha$ -KO d6 GFP+ cells against WT, and using heatmap2 to visualize the expression of these genes as the cells develop from ESCs towards d6 GFP+ cells. LC indicates luminal cell, EMP indicates embryonic multipotent progenitor cell, and BC indicates basal cell. LC/EMP and BC/EMP indicate markers present in both cell populations. The general classification of each gene is provided to the right of the gene name. Red indicates upregulation and blue indicates downregulation. . . . .	84

20	<p><b>Validation of SGET-based AP2<math>\gamma</math>-KO ESC line S-KO.</b> A) Comparison of translated AP2<math>\gamma</math> protein sequences between SGET-WT ESCs and AP2<math>\gamma</math>-KO ESC line S-KO. Premature stop codons are indicated as asterisks (*), and blue-highlighting demonstrates consensus between the WT and AP2<math>\gamma</math>-KO aa sequences. B) Western blot of SGET WT, S-Het, and S-KO ESC lysates, with 80 <math>\mu</math>g protein/lane. AP2<math>\gamma</math> isoforms 1 and 2 are indicated, with expected size in kDa of 49 and 53 kDa, respectively. The upper blot indicates <math>\alpha'</math>AP2<math>\gamma</math>-HRP labeling, while the lower blot indicates sequential <math>\alpha'</math>GAPDH-HRP labeling of the same blot, with expected size of 37 kDa. A protein ladder indicates sizes in kDa in the leftmost lane. C) 40x confocal images depicting the IF staining of SGET WT and S-KO ESCs. <math>\alpha</math>-AP2<math>\gamma</math>-Alexa-647 is shown in red and DAPI is shown in blue, with scale bars of 100 <math>\mu</math>m indicated. . . . .</p>	86
21	<p><b>KO of AP2<math>\gamma</math> impairs PGC marker expression on d6 of PGCLC differentiation in S-KO cells.</b> WT SGET and S-KO EpiLCs were aggregated as EBs for six days in either -cyto control medium or +cyto medium which induces the cells towards PGCLC fate. The EBs were subsequently fixed, cryosectioned, and labeled with <math>\alpha'</math>GFP-Alexa-488 (shown in green), <math>\alpha'</math>DAZL-Alexa-647 (shown in red), and DAPI. Overlay images of the <math>\alpha'</math>DAZL and <math>\alpha'</math>GFP channels are shown, as are scale bars indicating 50 <math>\mu</math>m. . . . .</p>	87
22	<p><b>Loss of GFP+/tdTomato- cells in the d6 +cyto EBs S-KO line indicates impaired PGCLC differentiation.</b> The percentage of GFP+/tdTomato- cells reporting Stella+/ESG1- expression in SGET WT and S-KO EBs grown in +/- cyto medium for six days is indicated. N = 3 for all samples. Significance (p &lt; 0.05) is indicated and was determined using the unpaired t-test in Prism 9. . . . .</p>	88
23	<p><b>S-KO d6 GFP+/tdTomato- cells do not upregulate PGC markers.</b> The qRT-PCR analysis of SGET EpiLC (grey bars), SGET d6 GFP+/tdTomato- cell (blue bars), and S-KO d6 GFP+/tdTomato- cell cDNA (red bars) for genes determined to be DE in d6 B1-based AP2<math>\gamma</math>-KO GFP+ cells is shown. Fold change is indicated relative to SGET-EpiLC, and standard deviation is indicated via error bars. Significance of p &lt; 0.05 was determined using the z-test in Microsoft Excel. N = 3 for all samples. . . . .</p>	89
24	<p><b>The TSS chromatin accessibility profile is not altered in AP2<math>\gamma</math>-KO cells.</b> The alignment of mapped reads at transcription start sites (TSS) of ATAC-seq libraries is shown as a heatmap, with one example of a WT or AP2<math>\gamma</math>-KO sample per cell type and KO status. The cell type and specific cell line used are both indicated, as are the number of reads in thousands (0-15), and the distance from the TSS in bp. Color indicates reads aligning to the TSS, with blue indicating high alignment and red indicating low alignment. The average TSS profile per library is shown above the heatmap. . . . .</p>	91
25	<p><b>Overlapping and unique peaks in ATAC-seq libraries.</b> The number of unique and overlapping peaks in each ATAC-seq library was determined using the HOMER function getDifferentialPeaks, and plotted using UpSetR. The number of total peaks per sample is indicated, as are the number of unique and shared peaks per individual sample or per multiple samples. A) ESC peaks. B) EpiLC peaks. C) D4 GFP+ peaks. D) D6 GFP+ peaks. . . . .</p>	92

26	<b>Interactive genome viewer (IGV) tracks of WT and AP2<math>\gamma</math>-KO ATAC-seq libraries with DAPs indicated at the <i>Tfap2c</i> gene locus.</b> WT1-3 and KO1-2 Bigwig files demonstrating genomically aligned reads are overlaid by cell type and AP2 $\gamma$ -KO status. Peak 1 indicates a peak that is differentially present in WT ESCs but is lost in AP2 $\gamma$ -KO ESCs and WT EpiLCs. Peak 2 indicates a peak that is present in all WT samples, however is reduced in size in AP2 $\gamma$ -KO ESCs and d4/d6 GFP+ cells. Chromosomal coordinates and increments are indicated. Transcription start sites for <i>Tfap2c</i> isoforms 1 and 2 are indicated in red adjacent to the chromosomal coordinates, with isoform 1 (indicated by “1”) located upstream of isoform 2 (indicated by “2”). . . .	93
27	<b>AP2<math>\gamma</math> and AP2<math>\alpha</math> are enriched in naïve- and PGC-specific ATAC-seq peaks.</b> Motif enrichment analysis of DAPs negatively or positively enriched in WT cells compared to the previous cell state. The rank and p-value calculated by HOMER are indicated, and significance is determined by p-value < 1e-3. A) Negatively enriched motifs in WT EpiLCs relative to ESCs. B) Positively enriched motifs in WT EpiLCs relative to ESCs. C) Negatively enriched motifs in d4 GFP+ PGCLCs relative to EpiLCs. D) Positively enriched motifs in d4 GFP+ PGCLCs relative to EpiLCs. . . . .	95
28	<b>Differential chromatin accessibility is greatest in AP2<math>\gamma</math>-KO ESCs.</b> MD plots depicting DAPs identified to be significant by EdgeR (FDR < 0.05, fc not considered; it should be noted that this significance is less stringent than that used for analysis in this work). Red indicates positive enrichment, while blue indicates negative enrichment. Log fold-change (LogFC) and average expression are indicated. MD plots represent: A) DAPs found between AP2 $\gamma$ -KO and WT ESCs, A) DAPs found between AP2 $\gamma$ -KO and WT EpiLCs, A) DAPs found between AP2 $\gamma$ -KO and WT d4 GFP+ cells, and D) DAPs found between AP2 $\gamma$ -KO and WT d6 GFP+ cells. . . . .	97
29	<b>AP2<math>\gamma</math> and AP2<math>\alpha</math> are negatively enriched in DAPs of AP2<math>\gamma</math>-KO ESCs.</b> Motif enrichment analysis of DAPs negatively or positively enriched in AP2 $\gamma$ -KO ESCs relative to WT. The rank and p-value calculated by HOMER are indicated, and significance is determined by p-value < 1e-3. A) Negatively enriched motifs. B) Positively enriched motifs. . . . .	99
30	<b>POU family motifs are positively enriched in DAPs of AP2<math>\gamma</math>-KO EpiLCs.</b> Motif enrichment analysis of DAPs negatively or positively enriched in AP2 $\gamma$ -KO EpiLCs relative to WT. The rank and p-value calculated by HOMER are indicated, and significance is determined by p-value < 1e-3. A) Negatively enriched motifs. B) Positively enriched motifs. . . . .	101
31	<b>AP2<math>\gamma</math> and AP2<math>\alpha</math> are negatively enriched in DAPs of AP2<math>\gamma</math>-KO d4 GFP+ cells.</b> Motif enrichment analysis of DAPs negatively or positively enriched in AP2 $\gamma$ -KO d4 GFP+ cells relative to WT. The rank and p-value calculated by HOMER are indicated, and significance is determined by p-value < 1e-3. A) Negatively enriched motifs. B) Positively enriched motifs. . . . .	102
32	<b>Only SOX9 and p73 are positively enriched in DAPs of AP2<math>\gamma</math>-KO d6 GFP+ cells.</b> Motif enrichment analysis of positively enriched DAPs in AP2 $\gamma$ -KO d6 GFP+ cells relative to WT. The rank and p-value calculated by HOMER are indicated, and significance is determined by p-value < 1e-3. . . . .	103

33	<b>Expression of neural fate genes increases with development towards PGCLC fate following KO of AP2<math>\gamma</math>.</b> Heatmap of genes associated with neural stem cells (NSCs), neural plate development, or neural crest fate. Red indicates increased expression, blue indicates downregulation. Significance (adj. p. < 0.01, fc +/- 1.5x) determined in AP2 $\gamma$ -KO d6 GFP+ cells against WT. Sample and gene names are indicated. . . . .	104
34	<b>Design and identification of plasmid pmTfap2c.</b> A) Plasmid pmTfap2c is comprised of the following features: the 5' <i>Tfap2c</i> homology arm (turquoise), a tdTomato cassette (red) separated from the <i>Tfap2c</i> 5' homology arm by a T2A site (pink), a PGK promoter (white) upstream of a puromycin resistance gene (green), a $\Delta$ TK gene (purple) immediately upstream of a bGH pol(A) signal (grey), the 3' <i>Tfap2c</i> homology arm (blue) separated from the poly(A), and an MC1-DTA cassette (grey). The selection cassette is flanked by two ROX sites (pink). B) pmTfap2c was Gibson-cloned and transformed into DH5 $\alpha$ transformation-competent bacteria. Single clones were isolated and resulting plasmids were purified and digested with SnaBI/XhoI to identify correct clones. Here, two bands indicate correct cloned pmTfap2c plasmids, with sized in kb/bps indicated. .	106
35	<b>The BGAT reporter line demonstrates reduced tdTomato expression following AP2<math>\gamma</math>-KO.</b> A) gDNA isolated from B1 and BGAT cells was genotyped for the presence of the WT <i>Tfap2c</i> allele, the tdTomato cassette, and four combinations of knock-in (KI) primers. No-template control was included, and sized in bps are indicated. B) pMiniT <sup>TM</sup> cloning and Sanger sequencing was translated to aa sequences in SnapGene <sup>®</sup> . Premature stop codons are indicated as asterisks (*), and blue-highlighting demonstrates consensus between the WT and AP2 $\gamma$ -KO aa sequences. C) 40x confocal images depicting the ICC staining of BGAT and BGAT-KO ESCs. $\alpha$ 'AP2 $\gamma$ -Alexa-488 is shown in green, $\alpha$ 'tdTomato-Alexa-647 is shown in purple, and DAPI is shown in blue, with scale bar of 100 $\mu$ m indicated. D) The percentage of tdTomato+ BGAT and BGAT-KO ESCs, potentially reporting AP2 $\gamma$ expression, is indicated. N = 3 for all samples. Significance (p < 0.05) is indicated, and was determined using the unpaired t-test in Prism 9. . . . .	107
36	<b>Expression of tdTomato in BGAT cells during PGCLC differentiation.</b> A) FACS profiles of B1 WT ESCs, BGAT ESCs, and BGAT EpiLCs with GFP and tdTomato fluorescence indicated. Gating was performed around B1 WT ESCs to indicate tdTomato+/GFP+ fluorescence in the daughter BGAT line. The overlay of BGAT ESC and EpiLC profiles is included, with the colored population indicating cell type - ESC in blue, EpiLC in red. B) FACS profiles of BGAT d6 -cyto and +cyto cell populations with GFP and tdTomato fluorescence indicated. Gating was performed around the d6 BGAT -cyto population to identify gained GFP+/tdTomato+ fluorescence upon the addition of cytokines, and the overlay of +/-cyto populations indicates -cyto cells in blue and +cyto cells in red. C) Statistical analysis of the percentage of BGAT d6 tdTomato+/GFP+ cells +/-cyto, with significance (p < 0.05) calculated via unpaired t-test in Prism 9 indicated. N = 3 for all samples. . . . .	109
37	<b>Unique and shared gRNAs per cell type and enrichment status.</b> Venn diagram depicting negatively and positively enriched gRNAs identified to be significant (FDR < 0.05) per cell type analyzed in screen. Overlapping regions without numbers do not indicate shared gRNAs, rather are a spatial necessity of the graph. Cell type and enrichment status are indicated in key. Venn diagram generated using DeepVenn. . . . .	112

38	<b><i>Tfap2c</i> TAD and surrounding regions demonstrate coverage of gRNA library and significant hits by cell type and enrichment status.</b> gRNA coverage of the <i>Tfap2c</i> TAD is indicated in the lane “gRNA library”. Significantly enriched (FDR < 0.05) gRNAs are identified in each indicated lane by cell type and positive or negative enrichment status. WT ATAC-seq and RNA-seq data is included for reference, with NGS library and cell type indicated. The <i>Tfap2c</i> gene is indicated and digitally enlarged to demonstrate significant gRNAs. Image generated in IGV with genes and chromosomal coordinates and increments indicated. . . . .	115
39	<b>Genomic features of significant gRNAs.</b> Pie charts of genomic features identified in significantly negatively- or positively-enriched gRNAs per cell type are indicated. A features key is included below, and “no features” indicates broadly intergenic sequences exhibiting no overlap with the listed features. . . . .	116
40	<b>The <i>Aurka/Cstf1</i> locus on chromosome 2 indicates negative enrichment of significant gRNAs in exons following EpiLC induction.</b> Significant gRNAs are plotted in IGV and demonstrate enrichment of exon-targeting gRNAs in the EpiLC negative enrichment sample. WT RNA-seq and ATAC-seq data are included and labeled for reference, and chromosomal coordinates and increments are noted. Image generated in IGV. . . . .	117
41	<b>Map of 2C::tdTomato reporter plasmid.</b> All features are noted, as well as RE sites used during the course of this work to linearize the plasmid. . . . .	xix
42	<b>Map of pHNANOS3-T2A-tdTomato-RoxPGKPuro plasmid.</b> All features are indicated and primers which were used during the course of this study to amplify features of the plasmid are indicated. . . . .	xx
43	<b>Map of pCAGGS-Dre-IH plasmid.</b> All features are indicated. . . . .	xxi
44	<b>Map of pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W plasmid.</b> All features and RE-digest sites used in this work are indicated. . . . .	xxii
45	<b>Map of pMD2.G plasmid.</b> All features are indicated. . . . .	xxiii
46	<b>Map of pmTfap2c plasmid.</b> All features, primers, RE cut-sites, and Gibson oligos relevant to this work are indicated. . . . .	xxiv
47	<b>Map of pX330-U6-Chimeric_BB-CBh-hSpCas9 plasmid.</b> All features are indicated, as well as the BbsI cut-sites relevant to this work. . . . .	xxvi
48	<b>Quality control of ATAC-seq libraries.</b> All libraries were aggregated per test with MultiQC. A) FastQC analysis: percentage of GC content of all ATAC-seq libraries pre- and post-Cutadapt-based adapter trimming. All but two samples “passed” as indicated by green or yellow curves; two samples “failed” as indicated by red curves. B) analysis: mean quality scores of all ATAC-seq libraries pre- and post-Cutadapt-based adapter trimming. All samples “passed” as indicated by the green line. C) SAMtools stats of HISAT2 alignment of all ATAC-seq libraries post-Cutadapt adapter-trimming. Blue indicates mapped, red indicates unmapped. D) Picard deduplication statistics of all ATAC-seq libraries post-Cutadapt adapter-trimming. Duplication status is indicated by color, and explained in the associated key. E) FastQC analysis: sequence length distribution of all ATAC-seq libraries pre- and post-Cutadapt-based adapter trimming. All pre-trimming samples passed, as indicated by the yellow line, and all post-trimming samples failed, as indicated by the green line. . . . .	xxvii

49	<b>Fragment length distribution of each ATAC-seq libraries used in this work.</b> Insert size in bps and counts in thousands are indicated. A) ESC WT and KO ATAC-seq libraries. B) EpiLC WT and KO ATAC-seq libraries. C) d4 GFP+ WT and KO ATAC-seq libraries. D) d6 GFP+ WT and KO ATAC-seq libraries. . . . .	xxviii
50	<b>Transcription start sites of ATAC-seq libraries.</b> The alignment of mapped reads at transcription start sites (TSS) of each ATAC-seq library is shown, with cell type and KO status indicated, as well as number of reads in thousands (0-15), and distance from TSS in bp are indicated. Color indicates number of reads aligning to the TSS, with blue indicating high alignment and red indicating poor alignment. The average TSS profile per library is shown. . . . .	xxix
51	<b>PCA of ATAC-seq libraries used in this work.</b> Percentage of variance in PC1 and PC2 are indicated. Clusters by cell type and/or KO status are indicated by black rings, key is also indicated. . . . .	xxx
52	<b>MD plots of DAPs identified between WT cell-state transitions.</b> Peaks significantly negatively or positively enriched are indicated, as is the average expression and logFC. A) DAPs enriched in WT EpiLCs relative wt WT ESCs. B) DAPs enriched in WT d4 GFP+ PGCLCs relative to WT EpiLCs. . . . .	xxx
53	<b>IGV tracks demonstrating DAPs identified in genes of interest.</b> IGV tracks demonstrate WT and AP2 $\gamma$ -KO overlaid Bigwig files representing accumulated reads. Chromosomal coordinates and increments are indicated. A) <i>Peg13</i> locus demonstrating DAP in indicated black box. B) <i>Mest</i> locus demonstrating DAP in indicated black box. C) <i>Pou5f1</i> locus demonstrating two DAPs in indicated black boxes. CREs identified by ENCODE are indicated in red adjacent to the chromosomal tracks. . . . .	xxxii
54	<b>Graphs indicating number of genes annotated to DAPs in ATAC-seq data.</b> The total number of genes annotated to DAPs (in grey), genes annotated to DAPs corresponding to DE genes in RNA-seq data (in green), genes annotated to DAPs corresponding to promoter regions (in blue), and genes annotated to DAPs corresponding to promoter regions of genes identified as DE in RNA-seq data (in purple) are shown by cell type and AP2 $\gamma$ -KO status. A) Genes annotated to DAPs in AP2 $\gamma$ -KO ESCs. B) Genes annotated to DAPs in AP2 $\gamma$ -KO EpiLCs. D) Genes annotated to DAPs in AP2 $\gamma$ -KO d4 GFP+ cells. D) Genes annotated to DAPs in AP2 $\gamma$ -KO d6 GFP+ cells. . . . .	xxxiii
55	<b><i>Tfap2c</i> tiling screen sequencing statistics and sample similarity information.</b> All plots were generated in MaGECK-VISPR, for all SGET ESC, EpiLC, and PGCLC libraries. A) Mean sequence quality of tiling screen libraries. B) Mean percentage GC content of tiling screen libraries. C) gRNA count in log10 of tiling screen libraries. D) Percentage mapped and unmapped reads of tiling screen libraries. E) Missed gRNA read counts in log10 of tiling screen libraries. F) Gini index of tiling screen libraries. G-I) PCA plots of principal components 1-3 of tiling screen libraries, principal components indicated per plot. J) Sample-to-sample distance plot of tiling screen libraries. . . . .	xxxiii
56	<b>Heatmaps of genes tiled with gRNAs in <i>Tfap2c</i> tiling screen.</b> Heatmap indicates expression of genes tiled in <i>Tfap2c</i> tiling screen over WT differentiation from ESC towards PGCLC fate. 1 indicates significant differential expression (adj. p. < 0.05) between ESC and EpiLCs, 2 indicates significant differential expression between EpiLCs and d4 PGCLCs. Red indicates upregulation, blue indicates downregulation. . . . .	xxxiv

57	<b>Quality controls of RNA-seq libraries.</b>	All libraries were aggregated per test with MultiQC. A) HISAT2 genomic alignment of libraries, with color indicating alignment status. Key included. B) Picard deduplication statistics of libraries, where color indicates duplication status. Key included. C) featureCounts assignments of libraries, where color indicates type of feature. Key included. D) SAMtools idxstats demonstrating fraction of total reads mapped per chromosome. Color indicates library, with outlier library WT1 EpiLC indicated in red text. E) RSeQC gene body coverage plot, demonstrating coverage of reads over gene body from 5' to 3' end, color indicates library. F) RSeQC infer experiment analysis of libraries, where color indicates sense of mapped strand. Key included. . . . .	xxxv
58	<b>Expression of SOX TF family genes increases with development towards PG-CLC fate following KO of AP2<math>\gamma</math>.</b>	Heatmap of SOX TF family genes. Red indicates increased expression, blue indicates downregulation. Significance was not determined. Sample and gene names are indicated. . . . .	xxxvii
59	<b>TEAD markers are differentially expressed in d6 GFP+ KO cells.</b>	Expression of each gene is indicated by normalized counts (log10 scale). Significance was determined between AP2 $\gamma$ -KO and WT samples of a given cell type using DESeq2. The color of each bar indicates the sample type, and is written in the key on the bottom right. N = 3 for WT samples, N = 2 for KO samples. . . . .	xxxviii
60	<b>Visualization of mapped RNA-seq reads on <i>Tfap2c</i> exon 2.</b>	Bigwig files generated from RNA-seq mapped reads are visually represented using IGV. WT1-3 and KO1-2 reads are overlaid by cell type, and are indicated to the left of each channel. The site at which KO3 resumes transcription of <i>Tfap2c</i> is indicated by the red bar above the reads and adjacent to the DNA distance scale bar. Chromosome 2 is indicated at the top of the scheme, and the red bar indicates the location of <i>Tfap2c</i> exon 2 within the chromosome. . . . .	xxxix

# List of Tables

1	DNA fragments required for the construction of plasmid pmTfap2c . . . . .	42
2	PGCLC control gRNAs found to be significant in sequenced cell types . . . . .	114
3	Requirements for the preparation of 5x ISO buffer . . . . .	i
4	Requirements for the preparation of Gibson assembly master mix: . . . . .	i
5	Requirements for the preparation of LB agar . . . . .	ii
6	Requirements for the preparation of 10x PBS . . . . .	ii
7	Requirements for the preparation of 10x TBE buffer . . . . .	ii
8	Antibodies used during the course of this work . . . . .	iii
9	N2B27 stock preparation . . . . .	iv
10	Requirements for the preparation of N2 supplement . . . . .	iv
11	Requirements for the preparation of N2B27 medium . . . . .	iv
12	Requirements for the preparation of 2i+LIF medium . . . . .	v
14	List of software and bioinformatics tools used in this work . . . . .	vii
15	Requirements for the preparation of a 10% SDS-PAGE gel . . . . .	viii
16	Requirements for the preparation of a 10% TBE gel . . . . .	viii
17	Disposables used in cell culture during this work . . . . .	ix
18	Media and reagents used in cell culture during this work . . . . .	x
19	Non-cell culture chemicals and reagents used during this work . . . . .	xiii
20	Non-cell culture disposables used during this work . . . . .	xv
21	Enzymes and mastermixes used during this work . . . . .	xvi
22	Equipment used during this work . . . . .	xvii
23	Kits used during this work . . . . .	xviii
24	<b>Map of psPAX2 plasmid.</b> All features are indicated. . . . .	xxv
25	CRISPR/Cas9 knockout gRNAs designed during this work . . . . .	xl
26	pmTfap2c Gibson assembly oligos designed during this work . . . . .	xli
27	Genotyping primers designed during this work . . . . .	xlii
28	Sequencing primers used during this work . . . . .	xlvi
29	qPCR primers used and designed during this work . . . . .	xlx
30	Genomic regions and genes targeted during the <i>Tfap2c</i> tiling screen . . . . .	li
13	Cell lines used during the course of this study . . . . .	lxv



**Table 13:** Cell lines used during the course of this study

Full name of cell line:	Abbreviation:	Type of cell:	Source:
Blimp1-GFP mouse embryonic stem cell	B1	mESC	Ohinata <i>et al.</i> , 2015 [26]
Blimp1-GFP AP2 $\gamma$ -KO gRNA #3 clone #2 mouse embryonic stem cell	KO1	mESC	Cera McDonald
Blimp1-GFP AP2 $\gamma$ -KO gRNA #3 clone #3 mouse embryonic stem cell	KO2	mESC	Cera McDonald
Blimp1-GFP AP2 $\gamma$ -KO gRNA #3 clone #11 mouse embryonic stem cell	KO3	mESC	Cera McDonald
Blimp1-GFP AP2 $\gamma$ -tdTomato Cas9 mouse embryonic cell line	BGAT	mESC	Cera McDonald
Blimp1-GFP AP2 $\gamma$ -tdTomato Cas9 mouse AP2 $\gamma$ -KO gRNA #2 clone #14 embryonic cell line	BGAT-KO	mESC	Cera McDonald
E14 mouse embryonic stem cells	E14	mESC	Hooper <i>et al.</i> , 1987 [185]
Human embryonic kidney SV40 large T antigen	HEK-293T	Human immortalized embryonic cell line	Graham <i>et al.</i> , 1977 [325]
Stella-eGFP Esg1-tdTomato mouse embryonic stem cell	SGET	mESC	Hackett <i>et al.</i> , 2018 [172]
Stella-eGFP Esg1-tdTomato AP2 $\gamma$ -KO gRNA #2 clone #13 mouse embryonic stem cell	S-KO	mESC	Cera McDonald
Stella-eGFP Esg1-tdTomato mouse embryonic stem cell Cas9 clone #10	SGET-Cas9	mESC	Cera McDonald



# Glossary

2i+LIF	Two inhibitors plus leukemia inhibitory factor; also written as 2i/LIF; see Appendix 7.3.1
3D	Three-dimensional
A	Adenine
aa	Amino acid
ACH	Active chromatin hubs
<i>Ad</i>	Towards (Latin) - add dH <sub>2</sub> O to this final volume
Adj. p.	Adjusted p-value
AiD	Auxin inducible degenon
<i>Alpl</i>	Alkaline phosphatase liver/bone/kidney (gene; for protein see TNAP)
AP	Alkaline phosphatase
A - P	Anterior to posterior axis
APS	Ammonium persulfate
AP2	Activation enhancer-binding protein 2 family
AP2 $\alpha$	Activating enhancer-binding protein 2 alpha (protein; for gene see <i>Tfap2a</i> )
AP2 $\beta$	Activating enhancer-binding protein 2 beta (protein; for gene see <i>Tfap2b</i> )
AP2 $\gamma$	Activating enhancer-binding protein 2 gamma (protein; for gene see <i>Tfap2c</i> )
AP2 $\delta$	Activating enhancer-binding protein 2 delta (protein; for gene see <i>Tfap2d</i> )
AP2 $\epsilon$	Activating enhancer-binding protein 2 epsilon (protein; for gene see <i>Tfap2e</i> )
<i>Arbp</i> /ARBP	Attachment region binding protein (gene/protein)
<i>ASCL1</i> /ASCL1	Achaete-scute family BHLH transcription factor 1 (human gene/protein)
<i>Asz1</i> /ASZ1	Ankyrin repeat SAM and basic leucine zipper domain containing 1 (gene/protein)
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
AVE	Anterior visceral endoderm
B1	Blimp1-meGFP ESC line
BC	Basal cell
bFGF	Basic fibroblast growth factor
BGAT	Blimp1-eGFP AP2 $\gamma$ -tdTomato ESC line
BGAT	Blimp1-eGFP AP2 $\gamma$ -tdTomato AP2 $\gamma$ -KO ESC line
bGH poly(A)	Bovine growth hormone polyadenylation signal
BHLH	Basic helix loop helix domain
BLAT	BLAST-like alignment tool
BLIMP1	B-lymphocyte-induced maturation protein 1 (protein; see <i>Prdm1</i> for gene)
<i>Bmp2</i> /BMP2	Bone morphogenetic protein 2 (gene/protein)
<i>Bmp4</i> /BMP4	Bone morphogenetic protein 4 (gene/protein)
<i>Bmp8a</i> /BMP8A	Bone morphogenetic protein 8A (gene/protein)
<i>Bmp8b</i> /BMP8B	Bone morphogenetic protein 8B (gene/protein)
<i>Bmpr1a</i> /BMPRI1A	Bone morphogenetic protein receptor 1A (gene/protein)
<i>Bmpr1b</i> /BMPRI1B	Bone morphogenetic protein receptor 1B (gene/protein)
<i>Bmpr2</i> /BMPRI2	Bone morphogenetic protein receptor 2 (gene/protein)

bp	Base pair
BR	Broad-range
Brachyury	Brachyury protein (for gene, see <i>T</i> )
BRN1	Brain1 (protein; for gene seen <i>Pou3f3</i> )
C	Cytosine
Ca	Calcium
Cas9	CRISPR associated protein 9
Cas9-LSD1	Cas9 fused to lysine-specific demethylase 1A
<i>Cbfa2t2</i> /CBFA2T2	CBFA2/RUNX1 partner transcriptional co-repressor 2 (gene/protein)
cCRE	Candidate <i>cis</i> -regulatory elements
<i>Cd44</i> /CD44	12505 (gene/protein)
<i>Cdkn1c</i> /CDKN1C	Cyclin dependent kinase inhibitor 1C (gene/protein)
CDS	Coding DNA sequence
<i>Cdx2</i> /CDX2	Caudal Type Homeobox 2 (gene/Protein)
<i>C. elegans</i>	<i>Caenorhabditis elegans</i> ; nematode
<i>Cer1</i> /CER1	Cerberus 1 (gene/protein)
<i>Chd3</i> /CHD3	Chromodomain helicase DNA binding protein 3 (gene/protein)
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing
Cl	Chlorine
<i>Col4a2</i> /COL4A2	Collagen type IV alpha 2 chain (gene/protein)
CpG	5'-C-phosphate-G-3', indicates a linear nucleotide sequence of CG (does not indicate a base pair of C-G)
CRE	<i>Cis</i> -regulatory element
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CSH	Cold Spring Harbor Laboratory
Ct	Cycle threshold
CTCF	CCCTC-binding factor
CTCFL	CCCTC-binding factor like
ctSCAN-SMS	<i>Cis</i> - and <i>trans</i> -regulatory elements scanning through saturating mutagenesis and sequencing
CUT&RUN	Cleavage under targets and release using nuclease
D2	Day 2
D4	Day 4
D6	Day 6
D10	Dulbecco's modified Eagle's medium with 10% FBS
DAPI	4',6-diamidino-2-phenylindole
dATP	Deoxyadenosine triphosphate
<i>Dazl</i> /DAZL	Deleted in azoospermia like (gene/protein)
dCas9-KRAB	Nuclease-dead Cas9 Krüppel-associated box
dCTP	Deoxycytidine triphosphate
<i>Ddx4</i> /DDX4	DEAD box helicase 4 (gene/protein); also known as <i>Vasa</i>

DE	Differentially expressed
DEAD Box	Nine conserved sequences, the second of which contains the amino acid sequence D-E-A-D (asp-glu-ala-asp)
<i>De novo</i>	From the beginning; anew (Latin)
dGTP	Deoxycytidine triphosphate
dH <sub>2</sub> O	Distilled water
<i>Dlk1</i> /DLK1	Delta-like non-canonical Notch ligand 1 (gene/protein)
<i>Dkk1</i> /DKK1	Dickkopf-related protein 1 (gene/protein)
dm	Dermal mesenchyme
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i> ; fruit fly
DMEM/F12	Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12
DNA	Deoxyribonucleic acid
DNase-seq	DNase treatment followed by sequencing assay
<i>Dnmt3b</i> /DNMT3B	DNA Methyltransferase 3 Beta (gene/protein)
<i>Dnmt3l</i> /DNMT3L	DNA Methyltransferase 3 like (gene/protein)
dNTP	Deoxynucleoside triphosphate
Dox	Doxycycline
DP	Differential peaks
dTTP	Deoxythymidine triphosphate
Dox	Doxycycline
<i>Dppa3</i>	Developmental pluripotency associated 3 (gene, for protein see Stella)
<i>Dppa5a</i>	Developmental pluripotency associated 5a (gene, for protein see ESG1)
<i>D. rerio</i>	<i>Danio rerio</i> ; zebrafish
DSB	Double strand break
DTA	Diphtheria toxin A
DTT	Dithiothreitol
DVE	Distal visceral endoderm
EB	Embryoid body
Ec	Ectoderm
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic
<i>e.g.</i>	<i>Exempli gratia</i> (Latin) - for example
EGF	Epidermal growth factor
eGFP	Enhanced green fluorescent protein
EKLF	Erythroid Krüppel like factor (protein)
ENCODE	Encyclopedia of deoxyribonucleic acid elements
EGTA	Ethylene glycol-bis( $\beta$ -aminoethyl ether)-N,N,N',N'-tetraacetic acid
EMP	Embryonic multipotent progenitor
EmVE	Embryonic visceral endoderm
<i>Eomes</i> /EOMES	Eomesodermin (gene/protein)

EpiLC	Epiblast-like cell
EpiSC	Epiblast-derived stem cell
eRNA	Enhancer ribonucleic acid
ESC	Embryonic stem cell
ESG1	Embryonal stem cell specific gene 1 (protein, for gene see <i>Dppa5a</i> )
<i>Esrrb</i> /ESRRB	Estrogen related receptor beta (gene/protein)
<i>Et al.</i>	<i>Et alia</i> (Latin) - and others
ETS	E-twenty-six erythroblast transformation specific family
<i>Ets1</i> /ETS2	ETS proto-oncogene 1 (gene/protein)
<i>Ets2</i> /ETS2	ETS proto-oncogene 2 (gene/protein)
<i>Etv5</i> /ETV5	ets variant 5 (gene/protein)
ExE	Extraembryonic ectoderm
ExVE	Extraembryonic visceral endoderm
<i>Ex vivo</i>	Outside of the body (Latin)
E14	E14 mouse embryonic stem cells
E(#)	Embryonic day #
dsRed	Discosoma red
FBS	Fetal bovine serum
fc	Fold change
<i>Fgf2</i> /FGF2	Fibroblast growth factor 2 (gene/protein)
<i>Fgf5</i> /FGF5	Fibroblast growth factor 5 (gene/protein)
<i>FIGLA</i> /FIGLA	folliculogenesis specific BHLH tran- scription factor (human gene/protein)
fMaSC	Fetal mammary stem cell
<i>Foxa1</i> /FOXA1	Forkhead box A1 (gene/protein)
FPP	Fat pad precursor
Fragilis	Fragilis protein (for gene see <i>Ifitm3</i> )
g	Gram
G	Guanine
G2 phase	Growth 2 phase
GATA	Nucleotide sequence - GATA proteins have affinity to bind to this nucleotide sequence
<i>Gata3</i> /GATA3	GATA Binding Protein 3 (gene/protein)
<i>Gata6</i> /GATA6	GATA Binding Protein 6 (gene/protein)
gDNA	Genomic DNA
GeCKO	Genome-scale clustered regularly interspaced short palindromic repeats knock-out
Germ plasm	Germinal cytoplasm
<i>Gfp</i> /GFP	Green fluorescent protein (gene/protein)
GMEM	Glasgow Modified Essential Medium
<i>Gnas</i> /GNAS	Guanine nucleotide binding protein, alpha stimulating (gene/protein)
<i>Grb10</i> /GRB10	Growth factor receptor bound 10 (gene/protein)
gRNA	Guide RNA

GSK3	Glycogen synthase kinase 3
<i>H1f0</i> /H1F0	Histone h1.0 linker protein (gene/protein)
H2A/H4R3me2	Symmetrical dimethylation of arginine 3 on histone 2A and histone 4
H3K4me1	Histone 3 lysine 4 monomethylation
H3K4me3	Histone 3 lysine 4 trimethylation
H3K9me2	Histone 3 lysine 9 dimethylation
H3K27ac	Histone 3 lysine 27 acetylation
H3K27me3	Histone 3 lysine 27 trimethylation
H4K20me1	Histone 4 lysine 20 monomethylation
H	Hydrogen
HCl	Hydrogen chloride
HDR	Homology directed repair
HEK-293T	Human embryonic kidney 293T cells
HeLa	Henrietta Lacks cell
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hESC	Human embryonic stem cell
HF	High fidelity
Hi-C	Hi-C chromatin conformation capture
<i>Hira</i> /HIRA	Histone cell cycle regulation (gene/protein)
<i>Hoxa1</i> /HOXA1	Homeobox A1 (gene/protein)
<i>Hoxa2</i> /HOXA2	Homeobox A2 (gene/protein)
<i>Hoxb1</i> /HOXB1	Homeobox B1 (gene/protein)
HRP	Horseradish peroxidase
<i>Ifitm3</i>	Interferon induced transmembrane protein 3 (gene; for protein see <i>Fragilis</i> )
HS	High-sensitivity
ICC	Immunocytochemistry
ICM	Inner cell mass
<i>Id</i>	Inhibitors of DNA binding gene family
<i>Id2</i> /ID2	Inhibitor Of DNA Binding 2 (gene/protein)
<i>i.e.</i>	<i>Id est</i> (Latin) - in other words
IF	Indirect immunofluorescence
<i>Ifitm3</i> /IFITM3	interferon induced transmembrane protein 3 (gene/protein); also known as <i>Fragilis</i> when translated
IgG	Immunoglobulin G
IGV	Interactive genome viewer
IHC	Immunohistochemistry
<i>In silico</i>	“In silicon” (pseudo-Latin); performed on the computer
<i>In vitro</i>	“In glass” (Latin) - experiment performed outside of natural biological context
<i>In vivo</i>	“Within the living” (Latin) - experiment performed within natural biological context
iPSC	Induced pluripotent stem cells
ISO	Isothermal

<i>Itgb3</i> /ITGB3	Integrin subunit beta 3 (gene/protein)
K	May refer to either potassium or lysine (context-dependent)
<i>Jam2</i> /JAM2	Junctional adhesion molecule 2 (gene/protein)
kb	Kilobase
KD	Knockdown
kDa	Kilodalton
KI	Knock-in
<i>Kit</i> /KIT	KIT proto-oncogene, receptor tyrosine kinase (gene/protein)
KLF	Krüppel like factor TF family
<i>Klf2</i> /KLF2	Krüppel Like Factor 2 (gene/protein)
<i>Klf4</i> /KLF4	Krüppel Like Factor 4 (gene/protein)
KO	Knockout
KO1	Blimp1-GFP AP2 $\gamma$ -KO gRNA #3 clone #2 mouse embryonic stem cell
KO2	Blimp1-GFP AP2 $\gamma$ -KO gRNA #3 clone #3 mouse embryonic stem cell
KO3	Blimp1-GFP AP2 $\gamma$ -KO gRNA #3 clone #11 mouse embryonic stem cell
<i>Krt8</i> /KRT8	Keratin 8 (gene/protein)
KSR	Knockout serum replacement
kV	Kilovolt
L	Liter
LC	Luminal cell
<i>Lefty1</i> /LEFTY1	Leftright determination factor 1 (gene/protein)
<i>Lefty2</i> /LEFTY2	Leftright determination factor 2 (gene/protein)
LogFC	Log fold-change
LIF	Leukemia inhibitory factor
LLC	Limited liability company
lncRNA	Long non-coding RNA
Ltd.	Limited company
M	Molar
mA	Milliamperes
MaGECK	Model-based analysis of genome-wide CRISPR/Cas9 knockout
MA plot	Log ratio (M) and mean average (A) plot
MaSC	Mammary stem cell
Mb	Megabase
MC1-DTA	Diphtheria toxin A under the control of the MC1 promoter
MEF	Mouse embryonic fibroblast
meGFP	Membrane-bound enhanced GFP
mESC	Mouse embryonic stem cell
<i>Meis2</i> /MEIS2	Meis homeobox 2 (gene/protein)
<i>Meis3</i> /MEIS3	Meis homeobox 3 (gene/protein)
MEK	Mitogen-activated extracellular signal-regulated kinase



mEpiLC	Mouse epiblast-like cell
mESC	Mouse embryonic stem cell; also referred to as mES
<i>Mest</i> /MEST	Mesoderm specific transcript (gene/protein)
Mg	Magnesium
MGI	Mouse genome informatics
Mio	Million
miRNA	Micro RNA
ml	Milliliter
MLE	Maximum likelihood estimation
mm	Mammary mesenchyme
mM	Millimolar
<i>M. musculus</i>	<i>Mus musculus</i> ; house mouse
mm10	GRCm38 mouse genome assembly
Mn	Manganese
mPGC	Mouse primordial germ cell
mPGCLC	Mouse primordial germ cell-like cell
mRNA	Messenger RNA
<i>Msx1</i> /MSX1	Homeobox protein Msx 1 (gene/protein)
<i>Msx2</i> /MSX2	Homeobox protein Msx 2 (gene/protein)
MOI	Multiplicity of infection
<i>Myc</i>	<i>Myelocytomatosis oncogene</i> (gene)
N	Any nucleobase
Na	Sodium
N/A	Not available
NAD <sup>+</sup>	Nicotinamide adenine dinucleotide
<i>Nanos3</i> /NANOS3	Nanos C2HC-Type Zinc Finger 3 (gene/protein)
NCE2	Negative control enhancer 2
n.d.	Not dated
NEAA	Non-essential amino acids
neg.	Negative, negatively
NFR	Nucleosome-free region
NFY	Nuclear transcription factor Y (protein; also known as NF-Y)
ng	Nanogram
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NMD	Nonsense-mediated decay
No.	Number
<i>Nodal</i> /NODAL	Nodal growth differentiation factor (gene/protein)
<i>Nr5a2</i> /NR5A2	Nuclear receptor subfamily 5 group A member 2 (gene/protein)
ns	Nipple sheath
NSC	Neural stem cell

O	Oxygen
OCT4	Octamer-binding transcription factor 4 (protein; for gene see <i>Pou5f1</i> )
OCT6	Octamer-binding transcription factor 6 (protein; for gene see <i>Pou3f1</i> )
<i>Otx2</i> /OTX2	Orthodenticle homeobox 2
P19	P19 embryonal carcinoma cells
P	Phosphorus
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif
PBS	Phosphate-buffered saline
<i>Pbx1</i> /PBX1	PBX homeobox 1 (gene/protein)
PC	Principal component
PCA	Principal component analysis
PCE3	Positive control enhancer 3
PCR	Polymerase chain reaction
P - D	Proximal to distal axis
PE	Parietal endoderm
<i>Pdgfra</i> /PDGFRA	Platelet-derived growth factor receptor A (gene/protein)
<i>Peg3</i> /PEG3	Paternally expressed 3 (gene/protein)
<i>Peg13</i> /PEG13	Paternally expressed 13 (gene/protein)
PEG 8000	Polyethylene glycol (molecular weight) 8000
PGC	Primordial germ cell
PGCLC	Primordial germ cell-like cell
PGK	Mouse phosphoglycerate kinase 1 promoter
pH	Potential of hydrogen
<i>PITX2</i> /PITX2	Paired like homeodomain 2 (human gene/protein)
pKLV2-W	pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W
pmTfap2c	pmTfap2c-T2A-tdTomato-RoxPGKPuro
Pol II	RNA polymerase II
Poly(A)	Polyadenylation
pos.	Positive, positively
POU class	Pituitary-specific Pit-1, Octamer transcription factor proteins Oct-1 and Oct-2, and Unc-86 transcription factor class proteins
<i>Pou3f1</i>	POU Class 3 Homeobox 1 (gene; see OCT6 for protein)
<i>Pou3f3</i>	POU Class 3 Homeobox 3 (gene; see BRN1 for protein)
<i>Pou5f1</i>	POU Class 5 Homeobox 1 (gene; see OCT4 for protein)
<i>Ppp4r1l-ps</i>	Protein phosphatase 4, regulatory subunit 1-like, pseudogene
PPxY	Proline-rich motif
<i>Prdm1</i>	PR/SET Domain 1 (gene; see BLIMP1 for protein)
<i>Prdm14</i> /PRDM14	PR/SET Domain 14 (gene/protein)
PR domain	PRDI-BF1-RIZ1 homologous region
PrE	Primitive endoderm

<i>Prm1</i> /PRM1	Protamine 1 (gene/protein)
PRR1	Potential regulatory region 1
PS	Primitive streak
P/S	Penicillin/Streptomycin
Puro	Puromycin
PVDF	Polyvinylidene fluoride
pX330	pX330-U6-Chimeric_BB-CBh-hSpCas9
qPCR	Quantitative polymerase chain reaction
QC	Quality control
qRT-PCR	Quantitative reverse-transcription polymerase chain reaction
R	Nucleotide which may be adenine or guanine
RA	Retinoic acid
RE	Restriction enzyme
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
RT	Room temperature
rtTA	Reverse tetracycline-controlled transactivator
S	Nucleotide that may be either guanine or cytosine
SCF	Stem cell factor
scRNA-seq	Single cell RNA-seq
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
<i>Serpinh1</i> /SERPINH1	Serpin family H, member 1 (gene/protein)
SET domain	Su(var), E(z), and Trithorax domain
SGET	Stella-eGFP Esg1-tdTomato ESC cell line
SGET-Cas9	SGET ESC line with genomically inserted Cas9 transgene
S-Het	Uncharacterized putative heterozygous AP2 $\gamma$ -KO SGET ESC line
<i>Shh</i>	Sonic hedgehog (gene)
S-KO	Stella-eGFP Esg1-tdTomato AP2 $\gamma$ -KO gRNA #2 clone #13 mouse ESC line
SMAD	Suppressor of mothers against decapentaplegic
<i>Smad1</i> /SMAD2	Suppressor of mothers against decapentaplegic 1 (gene/protein)
<i>Sma42</i> /SMAD2	Suppressor of mothers against decapentaplegic 4 (gene/protein)
<i>Smad5</i> /SMAD2	Suppressor of mothers against decapentaplegic 5 (gene/protein)
<i>Smad8</i> /SMAD2	Suppressor of mothers against decapentaplegic 8 (gene/protein)
<i>Sox1</i> /SOX1	SRY (sex determining region Y)-box 1 (gene/protein)
<i>Sox2</i> /SOX2	SRY (sex determining region Y)-box 2 (gene/protein)
<i>Sox3</i> /SOX3	SRY (sex determining region Y)-box 3 (gene/protein)
<i>Sox2</i> /SOX2	SRY (sex determining region Y)-box 4 (gene/protein)
<i>Sox4</i> /SOX4	SRY (sex determining region Y)-box 9 (gene/protein)

<i>Sox10</i> /SOX10	SRY (sex determining region Y)-box 10 (gene/protein)
<i>Sox11</i> /SOX11	SRY (sex determining region Y)-box 11 (gene/protein)
<i>Sox13</i> /SOX13	SRY (sex determining region Y)-box 13 (gene/protein)
<i>Sox17</i> /SOX17	SRY (sex determining region Y)-box 17 (gene/protein)
<i>S. pyogenes</i>	<i>Streptomyces pyogenes</i>
SRY	Sex-determining region Y
<i>Stra8</i> /STRA8	Stimulated by retinoic acid gene 8 (gene/protein)
SUMO	Small ubiquitin-like modifier
T2A	Thosea asigna virus 2A sequence
T	Thymine
<i>T</i>	<i>T</i> gene (for protein, see Brachyury)
TAD	Topologically associating domain
TAP-seq	Targeted perturb-sequencing
Taq	<i>Thermus aquaticus</i> polymerase
<i>Tdtomato</i> /tdTomato	Tandem dimer tomato (gene/protein)
TE	Trophectoderm
TE buffer	Tris-EDTA buffer
TEA domain	Nucleotide sequence CATTCCA/T
TEAD	TEA domain TF family
<i>Tead4</i>	TEA Domain Transcription Factor 4
TEB	Terminal end bud
TEMED	Tetramethylethylenediamine
TET	Ten-eleven translocation
<i>Tet1</i> /TET1	Tet methylcytosine dioxygenase 1 (gene/protein)
TF	Transcription factor
<i>Tfap2a</i>	Transcription factor AP2 alpha (gene, for protein see AP2 $\alpha$ )
<i>Tfap2b</i>	Transcription factor AP2 beta (gene, for protein see AP2 $\gamma$ )
<i>Tfap2c</i>	Transcription factor AP2 gamma (gene, for protein see AP2 $\gamma$ )
<i>Tfap2d</i>	Transcription factor AP2 delta (gene, for protein see AP2 $\gamma$ )
<i>Tfap2e</i>	Transcription factor AP2 epsilon (gene, for protein see AP2 $\gamma$ )
TGC	Trophoblast giant cell
TJ	Tight junction
TNAP	Tissue nonspecific alkaline phosphatase (protein; for gene see <i>Alpl</i> )
tracrRNA	<i>Trans</i> -activating CRISPR RNA
TRE3G	Tet-On 3G
TSC	Trophoblast stem cell
TSS	Transcription start site
Tris	Tris(hydroxymethyl)aminomethane
UCSC	University of California Santa Cruz
UK	United Kingdom

USA	United States of America
V	Volts
VE	Visceral endoderm
<i>Wnt3</i> /WNT3	Proto-oncogene protein Wnt-3 (gene/protein)
WT	Wild-type
WW domain	Domain containing two-tryptophan (WW) conserved motif
<i>Wwox</i> /WWOX	WW domain-containing oxidoreductase (gene/protein)
X	Times concentrated or X-chromosome (context-dependent)
<i>Xist</i>	Xi-specific transcript
<i>Zfp42</i> /ZFP42	Zinc finger protein 42 (gene/protein)
ZGA	Zygotic genome activation
ZIC	Zinc finger of the cerebellum TF family
$\alpha$	Alpha (Greek); Against, targeting (English)
$\Delta$	Delta (Greek); Indicates deletion of genomic region (see $\Delta$ TK)
$\Delta$ Ct	Ct value of gene of interest relative to housekeeping gene (see Section 2.4)
$\Delta\Delta$ Ct	Ct value of sample of interest relative to control sample after normalization to housekeeping gene (see Section 2.4)
$\Delta$ TK	Tetramethylethylenediamine
$\mu$ f	Microfarad
$\mu$ l	Microliter
$\mu$ M	Micromolar
$\Omega$	Omega (Greek); Ohm, unit of electrical resistance
	Early stop codon
-cyto	GK15 medium without the addition of cytokines
+cyto	GK15 medium with the addition of cytokines BMP4, BMP8a, SCF, EGF, and inhibitor LIF
-/-	Indicates homozygous knockout