UNIVERSITY OF HELSINKI

Functional analysis of alternative *TFAP2C* promoter usage in hESC-based trophoblast modelling

Master's Degree Programme in Genetics and Molecular Biosciences
Genetics and genomics
Master's thesis

Author:
Viljami Wuollet

Supervisor(s):
Jere Weltner, PhD

9.9.2026
Helsinki

Faculty: Faculty of Biological and Environmental Sciences

Degree programme: Master's Degree Programme in Genetics and Molecular Biosciences

Study track: Genetics and genomics

Author: Viljami Wuollet

Title: Functional analysis of alternative TFAP2C promoter usage in hESC-based trophoblast modelling

Level: Master's thesis

Month and year: September 2026

Number of pages: 70 + 10

Keywords: TFAP2C, CRISPR/Cas, CRISPRa, trophoblast, pluripotent stem cell, embryonic development, placenta

Supervisor: Jere Weltner

Where deposited: Helsinki University Library

Additional information:

Abstract:

TFAP2C is a key transcription factor involved in trophectoderm differentiation and placental development, and it is among the first transcription factors activated at the 4-cell stage of human embryogenesis. Impaired TFAP2C activity or placental dysfunction is associated with adverse pregnancy outcomes, including miscarriage and preeclampsia. Despite its central role in early lineage specification and in stem cell models of human development, the mechanisms by which TFAP2C regulates stem cell fate remain incompletely understood. Sequencing data from human embryos indicate that early TFAP2C expression appears to involve a poorly annotated upstream promoter, designated promoter 2 (P2), in addition to the canonical promoter 1 (P1). However, whether P2 functions as an independent regulator of TFAP2C expression remains experimentally unresolved.

This study investigated the contributions of P2 to TFAP2C expression and to trophoblast-like cell differentiation. Using CRISPR/Cas9 ribonucleoprotein complexes with in vitro-transcribed sgRNAs, cell lines with biallelic deletions targeting either the canonical promoter/exon 1 or the P2-associated region containing the alternative exon 1 were generated. The P2-targeted deletion removed the alternative exon 1 and part of the P2 region but did not delete P2 in its entirety. In parallel, inducible CRISPR activation (CRISPRa) lines were established with P1- or P2-specific gRNAs, enabling selective activation of each promoter in *TFAP2C* wild-type and promoter-knockout backgrounds. TFAP2C expression and downstream responses were quantified by immunocytochemistry and RT-qPCR.

Selective activation of P2 was sufficient to induce robust TFAP2C protein expression. Importantly, P2 activation also restored TFAP2C expression in P1-deleted cells, demonstrating that under CRISPRa conditions, P2-directed activation can drive TFAP2C independently of P1 and compensate for the loss of P1 in this system. These findings identify TFAP2C P2 as a previously unrecognised regulatory element in human placental development with potential relevance for future therapies targeting placenta-related disorders.

Tiedekunta: Bio- ja ympäristötieteellinen tiedekunta

Koulutusohjelma: Genetiikan ja molekulaaristen biotieteiden maisteriohjelma

Opintosuunta: Genetiikka ja genomiikka

Tekijä: Viljami Wuollet

Otsikko: Vaihtoehtoisen *TFAP2C*-promoottorin käytön toiminnallinen analyysi hESC-pohjaisessa trofoblastimallinnuksessa

Työn laji: Maisterintutkielma

Kuukausi ja vuosi: Syyskuu 2026

Sivumäärä: 70 + 10

Avainsanat: TFAP2C, CRISPR/Cas, CRISPRa, trofoblasti, pluripotentit kantasolut, alkionkehitys, istukka

Ohjaaja tai ohjaajat: Jere Weltner

Säilytyspaikka: Helsingin yliopiston kirjasto

Muita tietoja:

Tiivistelmä:

TFAP2C on keskeinen transkriptiotekijä trofektodermin erilaistumisessa ja istukan kehityksessä. Se kuuluu ensimmäisiin transkriptiotekijöihin, jotka aktivoituvat ihmisen alkionkehityksen 4-soluvaiheessa. TFAP2C:n heikentynyt aktiivisuus ja istukan toimintahäiriöt ovat yhteydessä raskauden aikaisiin sairauksiin ja komplikaatioihin, kuten keskenmenoon ja preeklampsiaan. Huolimatta TFAP2C:n keskeisestä roolista varhaisessa solulinjojen määräytymisessä ja alkionkehityksen kantasolumalleissa, tarkkoja mekanismeja, joiden kautta TFAP2C:n ilmentymistä säädellään, ei vielä täysin tunneta. Varhaisalkiosta saatu sekvensointiaineisto osoittaa, että TFAP2C:n varhaista ilmentymistä näyttäisi ohjaavan annotoimaton promoottori (P2), kanonisen promoottorin 1 (P1) lisäksi. Kokeellisesti ei kuitenkaan ole vielä osoitettu, toimiiko P2 TFAP2C:n ilmentymisen itsenäisenä säätelijänä.

Tässä tutkielmassa selvitettiin P2:n vaikutusta TFAP2C:n ilmentymiseen ja trofoblastinkaltaisten solujen erilaistumiseen pluripotenteista kantasoluista. CRISPR/Cas9-menetelmällä tuotettiin solulinjoja, joissa P1-promoottorin ja eksoni 1:n alue tai P2:een liittyvä vaihtoehtoinen eksoni 1:n sisältävä alue oli poistettu bialleelisesti. P2-kohdennettu deleetio poisti vaihtoehtoisen eksonin 1:n ja osan P2-alueesta, mutta ei koko promoottoria. Samanaikaisesti muodostettiin indusoitavaa CRISPR-aktivaatiojärjestelmää (CRISPRa) ilmentäviä solulinjoja, joissa käytettiin P1:tä tai P2:ta kohdentavia gRNA:ita. Tämä mahdollisti kummankin promoottorin selektiivisen aktivoinnin sekä *TFAP2C*-villityypissä että P1- tai P2-poistetuissa soluissa. TFAP2C:n ja sen kohdegeenien ilmentymistä määritettiin immunosytokemian ja RT-qPCR:n avulla.

P2:n selektiivinen aktivointi oli yksinään riittävä indusoimaan voimakkaan TFAP2C-proteiinin ilmentymisen. Merkittävää oli, että P2-aktivointi aiheutti TFAP2C:n ilmentymisen myös soluissa, joista P1 oli poistettu. Tämä osoittaa, että P2 pystyy ohjaamaan TFAP2C:n ilmentymistä P1:stä riippumatta ja osittain kompensoimaan P1:n puuttumista. Tulokset osoittavat, että TFAP2C:n P2-promoottori on aiemmin tunnistamaton säätelykohta ihmisen varhaisalkion kehityksessä ja saattaa siten olla merkityksellinen istukan toimintahäiriöihin kohdistuvien tulevien hoitomuotojen kehittämisessä.

DEDICATION

For Raisa and my family, the constants
through every uncertainty.

“Sometimes the most interesting answers begin upstream.”
—Viljami Wuollet

Table of contents

Abbreviations 9

1 Introduction 10

1.1 The human placenta: development and function 10

1.2 TFAP2C in early development and trophoblast specification 13

1.2.1 Expression and developmental roles 13

1.2.2 TFAP2C as a pioneer-like factor in trophoblast development 13

1.3 Alternative promoter regulation of TFAP2C 14

1.3.1 Discovery of an upstream TFAP2C promoter at zygotic genome activation 14

1.3.2 Functional consequences of promoter-driven isoform switching 15

1.4 CRISPR/Cas9 gene editing and CRISPRa as tools for causal dissection of promoter function 17

1.4.1 CRISPR/Cas9 for promoter-specific knockout 17

1.4.2 CRISPR activation for promoter-specific gain-of-function 18

1.5 Objectives 19

2 Materials and Methods 20

2.1 Cell line and culture conditions 20

2.1.1 Cell line 20

2.1.2 Routine passaging 20

2.1.3 Cryopreservation and thawing 21

2.2 CRISPR/Cas9 KO guide RNA design and *in vitro* synthesis 21

2.2.1 KO gRNA target selection 21

2.2.2 Design and synthesis of gRNA DNA templates 23

2.2.3 PCR assembly of gRNA DNA templates 24

2.2.4 In vitro transcription of sgRNAs 25

2.3 Ribonucleoprotein (RNP) assembly and electroporation 26

2.3.1 Preparation of Cas9-sgRNA complexes 26

2.3.2 Cell preparation and electroporation 26

2.4 Clonal isolation of genome-edited hESC cells 27

2.5 Genomic DNA extraction and TFAP2C promoter knockout validation 27

2.5.1 Genomic DNA extraction 27

2.5.2 Knockout validation primer design 28

2.5.3 PCR reaction setup and cycling conditions 30

2.5.4 Sanger sequencing 32

2.6 Pluripotency qualification 32
2.6.1 Morphological assessment 33
2.6.2 Functional trilineage assessment 33
2.7 Promoter-specific activation of TFAP2C using CRISPRa and induction of trophoblast factor-expressing cells 35
2.7.1 Design of TFAP2C CRISPRa guide RNAs 35
2.7.2 CRISPRa vector construction 37
2.7.3 Establishment of CRISPRa cell lines 37
2.7.4 Antibiotic selection of CRISPRa guides 38
2.7.5 Induction of TFAP2C expression 38
2.8 Immunocytochemistry 38
2.9 Imaging and image processing of protein expression 39
2.10 Quantitative reverse transcription PCR (RT-qPCR) 39
2.11 Statistical testing of RT-qPCR results 41
2.12 Figure 2 data processing 42
2.13 Research ethics 42
2.13.1 Human embryonic stem cell research and genetic modification 42
2.13.2 Data protection and research-data management 43
2.13.3 AI use declaration 43
3 Results 44
3.1 Generation and molecular validation of TFAP2C promoter-specific knockout hESC lines 44
3.1.1 Design and bulk performance of CRISPR/Cas9 editing 44
3.1.2 Clonal isolation and screening 45
3.1.3 Nested PCR and Sanger sequencing confirm targeted promoter-region deletions 46
3.1.4 Pluripotency and trilineage potential of TFAP2C promoter knockout lines 48
3.2 Each TFAP2C promoter independently drives gene expression when activated by CRISPRa 49
3.2.1 TFAP2C protein expression requires the targeted promoter to be intact 50
3.2.2 Quantitative analysis confirms promoter specificity and detects weak cross-activation 51
3.3 Promoter-directed TFAP2C activation induces a GATA2/KRT7-associated trophoblast-like phenotype 54
3.3.1 TFAP2C CRISPRa generates GATA2+/KRT7+ epithelial-like colonies 54
4 Discussion 57

4.1 TFAP2C promoter 2 is a functional promoter 57

4.1.1 Promoter compensation under enforced TFAP2C activation may mask context-dependent promoter function 58

4.2 TFAP2C isoforms show quantitative differences in hESCs, but may diverge functionally *in vivo* 59

4.3 Promoter usage provides a mechanism for developmental diversification of TFAP2C function 61

4.4 Limitations and future directions 62

4.5 Conclusion 64

5 Data and code availability 65

6 Acknowledgements 66

7 Author contribution 67

8 References 68

9 Appendices 76

Appendix 1 Plasmid maps 76

Appendix 2 Additional data from alternative promoter 79

Appendix 3 Sanger sequences 80

Appendix 4 Immunofluorescence 80

Appendix 5 Key RT-qPCR test results 81

Appendix 6 Resource tables 83

Appendix 7 RT-qPCR Primer validation 85

Abbreviations

bp	Base pair
CAG	Promoter CMV enhancer, chicken Actin intron, rabbit Globin splice acceptor
Cas9	CRISPR-associated protein 9
cDNA	Complementary deoxyribonucleic acid
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
dCas9	Catalytically inactive Cas9, “dead” Cas9
DSB	Double-strand break
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EVT	Extravillous trophoblast
FBS	Fetal bovine serum
FGF	Fibroblast growth factor
gRNA	Guide RNA
HDR	Homology-directed repair
hESC	Human embryonic stem cell
KO	Knockout
lncRNA	Long noncoding RNA
mRNA	Messenger RNA
NHEJ	Non-homologous end joining
nt	Nucleotide
PAM	Protospacer adjacent motif
PB	PiggyBac transposon
PBS	Phosphate-buffered saline
RNP	Ribonucleoprotein
ROCK	p160 Rho-associated coiled-coil kinase
RT	Room temperature
RT-qPCR	Reverse transcription quantitative PCR
sgRNA	Single guide RNA
S. pyogenes	*Streptococcus pyogenes*
spCas9	*S. pyogenes* Cas9
TBE	Tris/borate/EDTA
tracrRNA	Trans-activating CRISPR RNA

1 Introduction

1.1 The human placenta: development and function

The human placenta is a transient organ that sustains normal fetal growth throughout pregnancy by mediating nutrient and gas exchange, hormone production, immune regulation, and waste removal (Chen et al. 2022; Derisoud et al. 2024). Its epithelial compartment arises from the trophectoderm and comprises three major trophoblast lineages. These cells invade maternal tissues, forming a hemochorial interface in which chorionic villi are directly bathed in maternal blood (Figure 1A). Impaired trophoblast development is a major contributor to complications, including miscarriage and preeclampsia, and early placental dysfunction can shape long-term health outcomes in both the mother and the offspring. Molecular mechanisms that govern trophoblast self-renewal and commitment are central to understanding placental disease (Gong et al. 2024; Roberts et al. 2009; Kim et al. 2024).

Cytotrophoblasts (CTBs) are proliferative progenitor cells that self-renew and generate two major differentiated trophoblast lineages (Okae et al. 2018). Through syncytial fusion, a subset of CTBs forms the syncytiotrophoblast (STB), a multinucleated outer epithelial layer of the placenta responsible for hormone production and maternal-fetal exchange (Figure 1A) (Yu et al. 2002). In parallel, other CTBs differentiate into extravillous trophoblasts (EVTs), which invade the decidua and remodel the uterine spiral arteries (Cartwright et al. 2002; Chen et al. 2022; Kim et al. 2023; Ohgushi et al. 2025).

Human trophoblast stem cells (hTSCs) and related organoid systems now provide tractable experimental models for studying placental disease mechanisms, particularly when access to first-trimester tissue is limited (Chen et al. 2022). Naïve human embryonic stem cells (hESCs) can directly generate hTSCs with bipotent differentiation capacity, providing a model that closely recapitulates early trophoblast development (Dong et al. 2020). Primed hESCs were once thought to lack this potential, but defined culture can convert them into trophoblast-like cells or stable TSCs that resemble first-trimester CTBs and can be expanded long-term (Dong et al. 2020; Ohgushi et al. 2025; Sokka et al. 2025; Pastor et al. 2018; Tietze et al. 2024; Jang et al. 2022).

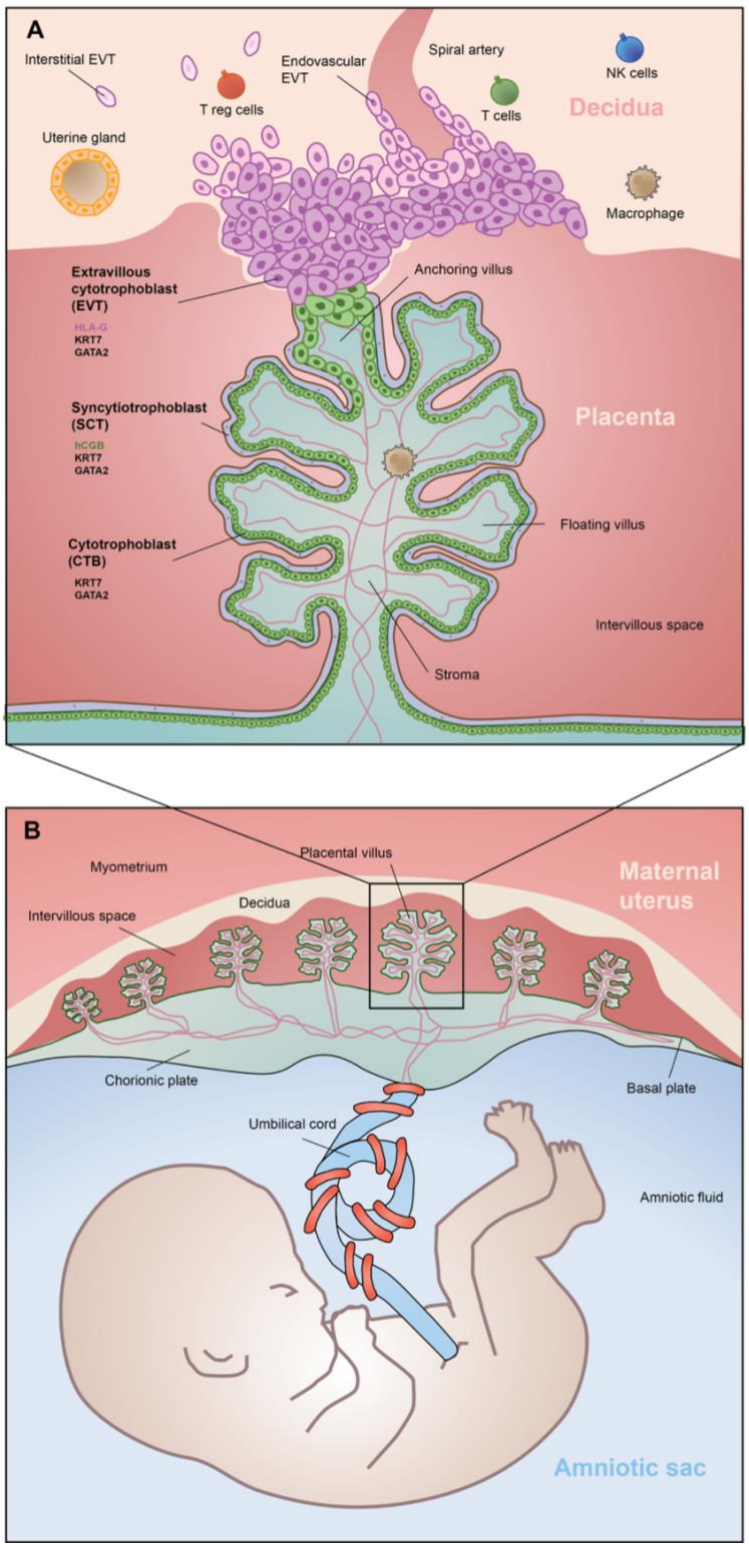


Figure 1. Structural organisation of the human placenta and trophoblast lineages. (A) Cellular architecture of the placental villus and maternal decidual interface. Anchoring villi penetrate the basal plate and generate extravillous cytotrophoblasts (EVTs), which invade the decidua, remodel spiral arteries, and interact with maternal immune cells, thereby establishing a controlled immunological and vascular environment. Floating villi are covered by a multinucleated syncytiotrophoblast (STB) layer that directly contacts maternal blood. Beneath the STB lies a layer of proliferative cytotrophoblasts (CTBs) that sustains villous growth and maintains epithelial integrity. **(B)** Overview of the fetal-maternal interface in late first- or second-trimester pregnancy. The fetus resides within the amniotic sac and is connected to the placenta via the umbilical cord. The chorionic villi extend from the chorionic plate into the maternal intervillous space. Maternal blood from the decidua bathes the villous surface, facilitating gas and nutrient exchange across the placental barrier. Illustration created by the author.

BMP was the first widely used inducer of trophoblast-like differentiation in human ESCs, driving cells through CTB-like intermediates towards STB and EVT characteristics (Li et al. 2013). Specification improves with ACTIVIN/NODAL inhibition, while adding WNT inhibitors reduces mesodermal contamination (Seetharam et al. 2022). BMP responses remain highly context-dependent, often yielding heterogeneous mixtures of trophoblast and mesodermal states rather than a uniform population. A major controversy concerns whether prolonged treatment of primed hESCs produces authentic trophoblasts or amnion-like cells (Roberts et al. 2018; Io et al. 2021).

Among the transcription factors that organise trophoblast development, TFAP2C has emerged as a key regulator of trophoblast identity, self-renewal, and lineage progression (Chen et al. 2022; Kim et al. 2023). TFAP2C often functions within enhancer-rich regulatory networks that sustain trophoblast-specific gene expression and stem-cell programmes (Krendl et al. 2017; Kim et al. 2023; Chen et al. 2022; Mak et al. 2021).

By maintaining trophoblast progenitor identity and coordinating lineage-specific transcriptional programmes, TFAP2C plays a critical role in guiding CTB differentiation towards the EVT lineage. TFAP2C promotes the tightly controlled transition from a proliferative state to an invasive phenotype, which is essential for proper uterine vascular remodelling. Deficiencies in this transcriptional regulation impair trophoblast invasion (Varberg et al. 2023; Kim et al. 2024). TFAP2C also appears to participate in STB differentiation at an earlier stage, likely by supporting progenitor competence and initial lineage specification rather than terminal syncytial maturation. TFAP2C expression decreases as villous trophoblasts differentiate, while TFAP2A becomes the predominant AP-2 family factor in mature STB cells (Kim et al. 2024; Handwerger 2010).

1.2 TFAP2C in early development and trophoblast specification

1.2.1 Expression and developmental roles

Transcription factor AP-2γ (TFAP2C) is an early regulator of cell-fate specification that links pluripotency-associated programmes to trophectoderm competence. In human embryos, TFAP2C expression increases markedly at the 8-cell stage, coinciding with major zygotic genome activation (ZGA), and remains detectable in all three blastocyst lineages (Wu et al. 2018). Mouse studies support a similarly early role, showing that Tfap2c contributes to outer-cell polarisation at the 8-cell stage (Zhu et al., 2020) and supports the transcriptional programmes of both the inner cell mass (ICM) and the trophectoderm (TE) during the totipotent period (Li et al., 2024). These findings support a model in which TFAP2C maintains a transient bipotent state from which blastomeres can initiate either embryonic or extraembryonic programmes.

1.2.2 TFAP2C as a pioneer-like factor in trophoblast development

TFAP2C functions as a pioneer-like transcription factor that helps establish trophoblast-specific enhancer landscapes. In both humans and mice, it is strongly enriched at enhancers associated with core trophoblast regulators, including CDX2, TEAD4, ELF5, and GATA2/3, while repressing competing embryonic differentiation programmes (Pastor et al. 2018). Loss of TFAP2C in trophoblast stem cells reduces stemness and impairs their differentiation potential, decreasing their capacity to produce invasive EVT populations (Ohgushi et al. 2025).

Conversely, overexpression of TFAP2C, particularly in combination with GATA2, GATA3, KLF5 and c-MYC, is sufficient to convert primed hESCs into trophoblast stem cell-like states (Balestrini et al. 2024). Although TFAP2A exhibits some functional overlap with TFAP2C in chromatin-opening capacity, comparative knockout and temporal binding studies indicate only partial redundancy. TFAP2A is more important for early polarisation and morula progression, whereas TFAP2C has broader roles in bipotency, trophoblast stemness, and invasive lineages (Liu et al. 2023). TFAP2C also directly regulates pluripotency. In human naïve pluripotent cells, it works with KLF factors and OCT4 to maintain naïve enhancer accessibility and is essential for the extraembryonic differentiation of naïve hESCs (Kumar et al. 2022). The balance between maintaining naïve identity and permitting trophoblast differentiation

highlights the need to study TFAP2C-specific mechanisms, such as alternative promoters, to understand how its pioneer activity is regulated across developmental stages.

1.3 Alternative promoter regulation of TFAP2C

1.3.1 Discovery of an upstream TFAP2C promoter at zygotic genome activation

Recent RNA-seq analyses of human preimplantation embryos have identified a previously unannotated transcription start site upstream of *TFAP2C,* hereafter referred to as promoter 2 (P2), located more than 4 kilobases upstream of the canonical promoter, P1 (Figure 2A) (Torre et al. 2023; Töhönen et al. 2015). These promoters produce alternative TFAP2C isoforms with distinct first exons that converge on the same downstream exons. Notably, this upstream region is currently annotated as a long non-coding RNA promoter rather than a TFAP2C promoter, creating an important annotation ambiguity. The locus may represent a misassigned promoter, a dual-use regulatory element, or a lncRNA-associated promoter that also contributes to *TFAP2C* transcription.

Promoter-specific transcription is strongly biased towards P2 during early cleavage stages. No promoter-specific TFAP2C is detected at the 2-cell stage, but P2-derived transcripts appear in 12 % of cells at the 4-cell stage. By the 8-cell stage, P2-derived transcripts are detected in 20 % of cells, compared with only 2 % for P1, while 78 % remain undetectable. This shift in promoter usage coincides with a sharp increase in overall TFAP2C expression (Figure 2C-D) during major zygotic genome activation (ZGA). Elevated TFAP2C expression is maintained through the morula and late blastocyst stages. Thus, the P2-derived isoform is the predominant detected promoter-specific *TFAP2C* transcript at the 8-cell stage, while the canonical P1 promoter retains a smaller contribution.

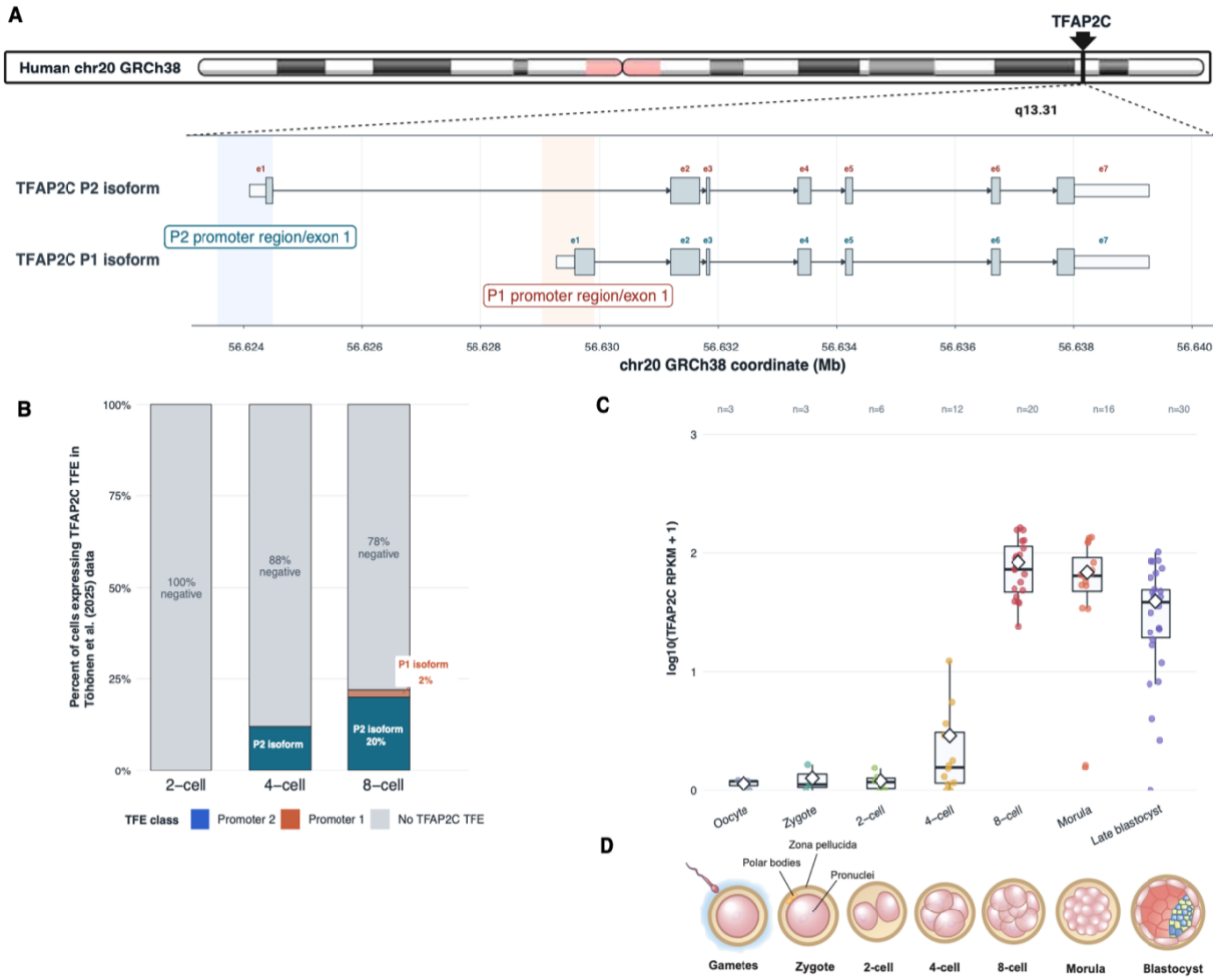


Figure 2. Early TFAP2C promoter usage and expression dynamics during human preimplantation development. (A) Genomic organization of the human TFAP2C locus on chromosome 20 (GRCh38), showing the canonical promoter P1 and upstream alternative promoter P2 and their corresponding transcript isoforms. The P1 and P2 isoforms contain distinct promoter-associated first exons but share the downstream *TFAP2C* exons. **(B)** *TFAP2C* promoter usage was inferred from the early human embryo STRT/TFE 5'-end transcription-initiation dataset of Töhönen et al. (2015), deposited in ENA/SRA BioProject PRJEB8994. *TFAP2C* TFE clusters were handled in GRCh38 coordinates and overlapped with the annotated P2 and P1 promoter/exon 1 regions. Cells were classified as P2-positive, P1-positive, or without detected *TFAP2C* TFE, and percentages were plotted per stage. **(C)** Total TFAP2C expression was plotted from the human preimplantation embryo single-cell RNA-seq dataset of Yan et al. (2013), deposited in GEO as GSE36552. RPKM expression values were grouped by developmental stage and transformed as log10(RPKM + 1), points show individual cells/samples, boxes show median and interquartile range, and diamonds show the mean. **(D)** Schematic overview of human preimplantation stages from fertilisation through the blastocyst stage, indicating the developmental stages corresponding to the expression and promoter-usage analyses.

1.3.2 Functional consequences of promoter-driven isoform switching

Alternative promoters can reprogram mRNA behaviour by altering 5'UTR structure, transcript stability, translational efficiency, and cell-type specificity (Ayoubi et al. 1996; Tan et al. 2025). Several key developmental regulators, including OCT4 and TP63, utilise alternative promoters or splicing to generate isoforms with distinct regulatory states during differentiation (Takeda et al. 1992; Murari et al. 2025).

As shown in Figure 3, transcripts from the P1 and P2 promoters differ in their first exons, producing N-terminally distinct protein isoforms. The P1-derived isoform contains an N-terminal sequence (MLWKITDNVKYEEDCE) that includes a SUMOylation motif at lysine 10 (Eloranta and Hurst, 2002). This post-translational modification influences transcriptional activity, cofactor recruitment, and responsiveness to upstream signalling pathways (Eloranta and Hurst, 2002). By contrast, the P2-derived isoform begins with MSFQFLR and lacks this motif. The P1-specific SUMOylation motif raises the possibility that the isoforms differ in post-translational regulation, however, this was not tested here.

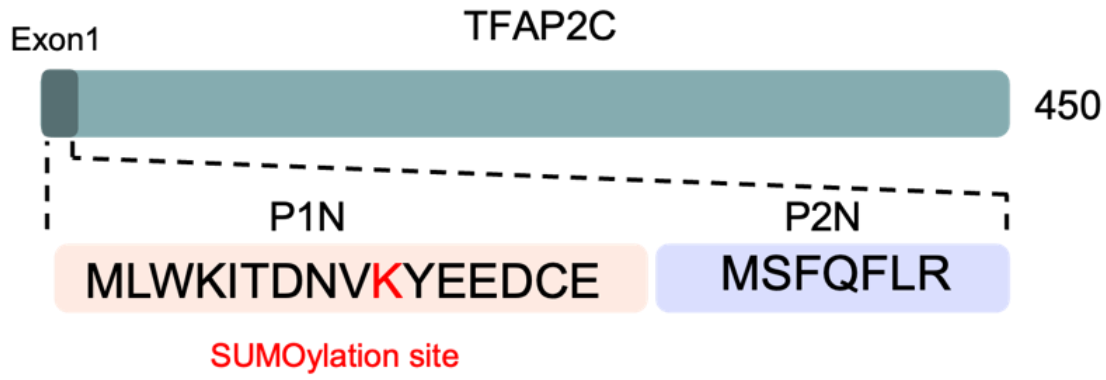


Figure 3. TFAP2C N-terminal isoforms generated by alternative promoter usage*. Schematic showing the distinct N-terminal sequences encoded by transcripts from the P1 and P2 promoters. The P1-derived isoform contains a SUMOylation motif at lysine 10 (highlighted in red), whereas the P2-derived isoform lacks this motif.*

These structural differences could contribute to context-dependent functional divergence at the cellular level. In line with studies showing that promoter choice can alter the regulatory states of key developmental factors (Ayoubi et al. 1996; Tan et al. 2025), the use of P2 versus P1 may control whether TFAP2C operates in a tightly constrained SUMOylation-regulated mode or a more permissive, activation-biased mode of gene regulation during trophoblast differentiation. This promoter-switching mechanism may provide a precise way to fine-tune trophoblast behaviour and placental development and offers a tool for disease modelling using isoform-specific perturbations *in vitro*.

1.4 CRISPR/Cas9 gene editing and CRISPRa as tools for causal dissection of promoter function

1.4.1 CRISPR/Cas9 for promoter-specific knockout

CRISPR/Cas9 is a programmable nuclease system derived from a bacterial adaptive immune mechanism that enables site-specific genome modification by introducing targeted DNA double-strand breaks (DSBs) (Figure 4). A single-guide RNA (sgRNA) directs the Cas9 endonuclease to a 20-nucleotide target sequence adjacent to an NGG protospacer-adjacent motif (PAM). The HNH and RuvC domains then cleave opposite DNA strands, generating a blunt DSB. Repair is mediated by endogenous DNA repair pathways. Most commonly, the error-prone non-homologous end joining (NHEJ) pathway ligates the break, often producing small insertions or deletions that disrupt the target sequence. Although less efficiently, homology-directed repair (HDR) can introduce precise sequence changes when a donor template is provided (Jinek et al. 2012; Hsu et al. 2014).

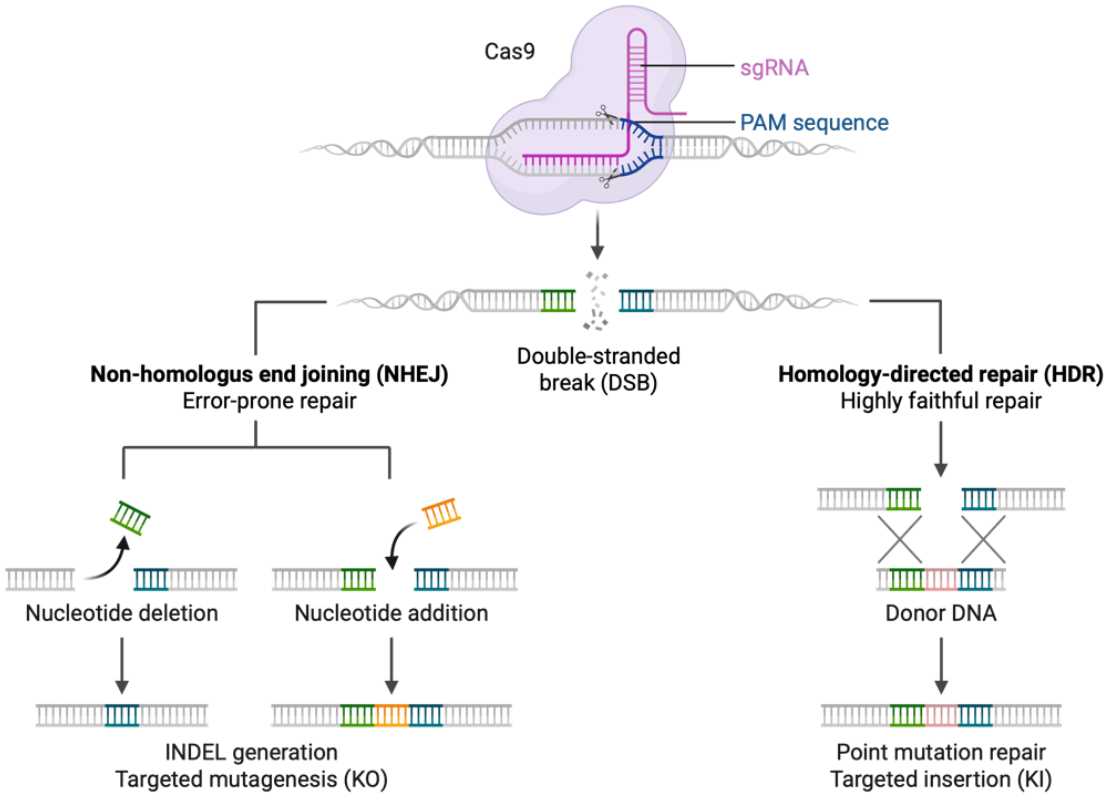


Figure 4. Mechanism of CRISPR/Cas9 genome editing. *Schematic overview of Cas9-directed DNA cleavage and subsequent repair outcomes. An sgRNA guides Cas9 to a genomic target adjacent to a PAM sequence, where Cas9 generates a double-strand break. Repair by non-homologous end-joining (NHEJ) produces insertions or deletions that can disrupt the target locus, while homology-directed repair (HDR) uses an external donor template to introduce precise sequence changes. Created with BioRender.com.*

In this study, paired sgRNAs were designed to flank each *TFAP2C* promoter and its corresponding isoform-specific exon 1, enabling Cas9 to excise the intervening segment and generate N-terminal variant-specific deletions. This approach enables selective disruption of transcription from one promoter while preserving the alternative promoter and the alternative TFAP2C isoform, in contrast to conventional whole-gene knockouts.

Because the upstream P2 promoter is currently annotated only as the transcription start site of a long non-coding RNA (lncRNA), these targeted deletions serve a critical validation role. Excising the P2-associated region containing alternative exon 1 allows direct testing of whether this region is *a bona fide* regulatory element driving TFAP2C expression, rather than a transcriptional artefact or a lncRNA promoter with incidental read-through. Furthermore, these alleles allow direct testing of this region for production of P2-derived transcript and reveal whether TFAP2C expression can be maintained through compensatory use of the remaining, unaffected promoter.

1.4.2 CRISPR activation for promoter-specific gain-of-function

CRISPR activation (CRISPRa) uses a catalytically inactive Cas9 (dCas9) fused to transcriptional activators such as VP64 or SAM to stimulate endogenous expression without inducing DSBs. Guide RNAs direct dCas9-activator complexes to promoter-proximal regions, where they recruit transcriptional machinery, remodel chromatin, and enhance transcriptional activation frequency (Omachi et al. 2022; Konermann et al. 2015; Gilbert et al. 2013). First-generation dCas9-VP64 produces only modest activation, whereas newer multicomponent systems such as SAM and SunTag achieve substantially stronger and more tunable upregulation (Omachi et al. 2022; Ding et al. 2022).

Here, an inducible dCas9-VP192 system, adapted from Balboa and colleagues, was used to promote promoter-specific *TFAP2C* transcription (Balboa et al. 2015). Targeting the activator complex to individual *TFAP2C* promoters in human pluripotent stem cells enables direct testing of whether each promoter can initiate transcription in its native chromatin context and of how selective activation influences transcriptional output and cell fate.

When combined with promoter-specific deletions, this dual-perturbation strategy assesses isoform necessity (via deletion) and sufficiency (via CRISPRa). Prior work shows that CRISPRa-mediated activation of single lineage-determining transcription factors, such as GATA3, ELF5, or TFAP2C, can efficiently drive pluripotent stem cells towards trophoblast fates, underscoring the power of CRISPRa for probing cell fate determinants (Krendl et al. 2017; Castel et al. 2020). Although *TFAP2C* has been extensively studied in both mouse and human systems, its regulatory architecture remains poorly understood.

This thesis addressed three main research questions:

(1) Can CRISPRa targeted to the upstream P2 region induce productive TFAP2C transcription and protein expression independently of P1?
(2) How do P1- and P2-associated exon 1 deletions affect promoter-resolved TFAP2C expression and the response to promoter-directed CRISPRa?
(3) Does activation through P1 versus P2 reproduce differences in the GATA2/KRT7-associated trophoblast-like phenotype under enforced TFAP2C activation?

1.5 Objectives

This thesis has four primary objectives:

(1) To generate H9 human embryonic stem cell (hESC) lines carrying either a TFAP2C P1-associated deletion affecting P1 and exon 1, or a partial P2-region deletion that removes the alternative exon 1 associated with the newly identified upstream promoter (P2), while retaining the upstream CRISPR activation target sites.

(2) To characterise these promoter-edited clones *in vitro* to determine whether they (a) maintain pluripotency and (b) retain the capacity to differentiate into trophoblast-like cells.

(3) To utilise an inducible CRISPR activation (CRISPRa) system to selectively activate each *TFAP2C* promoter within the respective deletion backgrounds, thereby assessing the ability of each isoform to drive trophoblast gene expression.

(4) To integrate these experimental findings with the current literature on TFAP2C function, promoter redundancy, and placental development, and to assess potential isoform-specific roles of TFAP2C in human stem cell differentiation towards the trophoblast lineage.

2 Materials and Methods

2.1 Cell line and culture conditions

2.1.1 Cell line

All experiments used H9 human embryonic stem cells (WA09, WiCell Research Institute; RRID: CVCL_9773) engineered to stably express a doxycycline-inducible CRISPR activation system (H9 SBact). The integrated cassette comprised a destabilised dCas9-VP192 fusion (DHFR-dCas9-VP192) under a TRE-tight promoter and a constitutively expressed tetracycline-responsive transactivator (M2-rtTA) under the CAG promoter. This configuration enabled dose-dependent transcriptional activation by doxycycline and additional regulation of dCas9 stability by trimethoprim (TMP) via the DHFR destabilisation domain.

H9 SBact cells were generated using the Sleeping Beauty (SB100X) transposon system by co-transfecting the SB transposase with donor plasmids encoding TRE::DHFR-dCas9-VP192 and CAG::rtTA, followed by dual selection with G418 and Zeocin (Gawriyski et al., 2023). Cells were maintained under feeder-free conditions on tissue culture plates coated with human recombinant laminin-521 (BioLamina, LN521-05, 5 µg/mL in PBS containing Mg2+ and Ca2+), incubated overnight at 4 °C. Cells were cultured in Essential 8 (E8, Thermo Fisher, A1517001) with daily medium changes at 37 °C, 5% CO2. Colony morphology was monitored by phase-contrast microscopy to confirm pluripotent morphology.

2.1.2 Routine passaging

Cells were passaged every 3-5 days at 75-80% confluence using Versene (0.48 mM EDTA in PBS, Gibco, Cat#15040-066). Cultures were rinsed with calcium- and magnesium-free PBS (Thermo Fisher Scientific, 14190250) and incubated in Versene at 37 °C for 3-4 minutes until colony edges loosened. Colonies were gently detached in E8 medium, collected in 15 mL Falcon tubes, and reseeded at 1:10 in E8 (2 mL per well in 6-well plates). Cultures were allowed to recover for at least 48 hours before experimental manipulation, and fresh medium was changed daily.

2.1.3 Cryopreservation and thawing

H9 SBact and derived knockout/CRISPRa lines were cryopreserved regularly to maintain early-passage stocks. After passaging, cell suspensions in E8 were mixed 1:1 with freezing medium (20% DMSO, 40% FBS, 40% E8). Aliquots of 1 mL were transferred into labelled cryovials and frozen at -80 °C in an isopropanol freezing container. After 12–24 hours, the cryovials were transferred to liquid nitrogen for long-term storage.

For thawing, vials were rapidly warmed and transferred to 15 mL Falcon tubes. DMSO was diluted with pre-warmed E8. Cells were pelleted at 300 x g for 5 minutes and resuspended in fresh E8 supplemented with 10 μM Y-27632. Cells were plated into laminin-521-coated wells (1 vial per well in a 6-well plate).

2.2 CRISPR/Cas9 KO guide RNA design and *in vitro* synthesis

2.2.1 KO gRNA target selection

Two TFAP2C regulatory regions were targeted to dissect promoter-specific functions. Promoter 1 (P1) corresponds to the canonical transcription start site (TSS) and exon 1. Promoter 2 (P2) lies upstream of an alternative exon 1, identified by transcription initiation peaks in human preimplantation embryos. Paired sgRNAs were designed to excise part of each regulatory region, along with its corresponding exon 1, enabling isoform-specific loss-of-function while preserving the alternative isoform.

As shown in Figure 5, KO gRNAs 4-5 flank P1 and its exon 1. For the P2 target, the upstream guide cuts between the TSS and the start codon within the P2 transcript's 5' UTR, while the downstream guide flanks the alternative exon 1. This deletes the alternative first exon without deleting P2 in its entirety. Because both isoforms splice to exon 2, deleting one first exon preserves transcription from the other. Flanking primers outside each targeted interval were used to verify deletions by PCR.

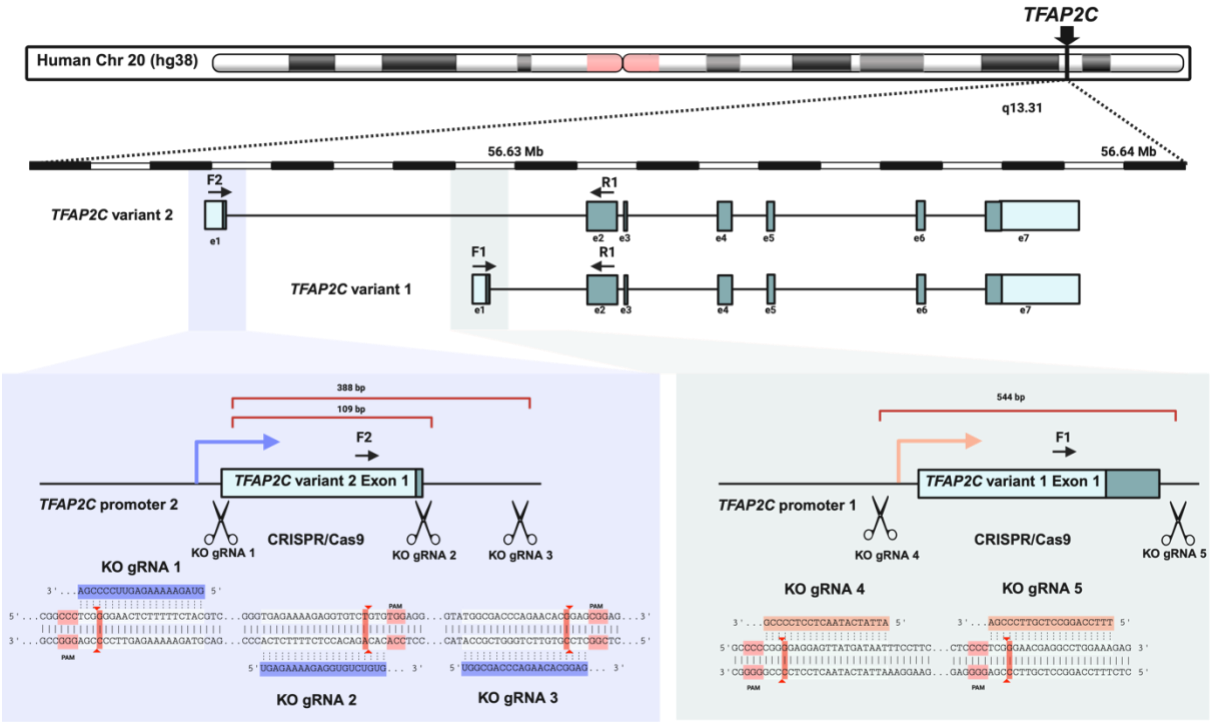


Figure 5. CRISPR/Cas9 design for TFAP2C promoter-specific deletions. *Paired gRNAs were designed to independently delete the promoter with exon 1 (KO gRNAs 4-5) or the alternative exon 1 associated with promoter 2 (KO gRNAs 1-3). Because both isoforms splice directly to exon 2, excision of one exon preserves transcription from the other, thus enabling a promoter-specific loss-of-function analysis. Flanking primers were designed outside each targeted interval to confirm deletions by PCR.*

Candidate KO gRNAs were designed using Benchling's CRISPR design platform. Predicted on-target activity was scored using the Doench et al. (2016) algorithm. Off-target potential was assessed using the MIT specificity score (Hsu et al., 2013). Guides were selected based on: (i) no predicted off-targets in GRCh38/hg38 with fewer than 3 mismatches, and (ii) 40-70% GC content.

For P1, KO gRNAs 4 and 5 were paired to delete the canonical promoter and exon 1. For P2, KO gRNA 1 was paired with gRNA2 or gRNA3 to excise a P2-associated region containing the alternative exon 1 (Fig. 5). Guide sequences, PAMs, genomic coordinates, and predicted scores are listed in Table 1.

Table 1. Properties of CRISPR/Cas9 gRNAs targeting *TFAP2C*. *Strand orientation is relative to the TFAP2C coding strand. PAM sequences correspond to NGG motifs recognised by spCas9. On-target scores (Doench et al. 2016) and off-target scores (Hsu et al. 2013) were obtained using Benchling's gRNA design tool in November 2024. Higher scores generally indicate better performance or specificity.*

Guide ID	Target region	Target sequence (5' -> 3')	Strand	PAM	Genomic locus (GRCh38)	On-target score	Off-target score
P1-KO-gRNA4	Promoter 1	ATTATCATAA CTCCTCCCCG	antisense	GGG	chr20:56629245 –56629265	78.7	44.3
P1-KO-gRNA5	Promoter 1	TTTCCAGGCCT CGTTCCCGA	antisense	GGG	chr20:56629789 –56629809	49.5	45.8
P2-KO-gRNA1	Promoter 2	GTAGAAAAAG AGTTCCCCGA	antisense	GGG	chr20:56625095 –56625115	75.0	43.7
P2-KO-gRNA2	Promoter 2	TGAGAAAAGA GGTGTCTGTG	sense	TGG	chr20:56625190 –56625210	74.4	30.6
P2-KO-gRNA3	Promoter 2	TGGCGACCCAG AACACGGAG	sense	NGG	chr20:56625469 –56625489	70.0	44.1

2.2.2 Design and synthesis of gRNA DNA templates

For in vitro transcription (IVT), gRNA DNA templates were assembled in a modular format (Fig. 6, Table 2). Each template comprised three elements: (i) an 18-nt T7 promoter, (ii) a 20-nt target-specific sequence immediately upstream of the PAM, and (iii) an 80-nt crRNA/tracrRNA scaffold for Cas9 binding. This design enabled rapid exchange of spacer sequences while maintaining a common scaffold for all gRNAs.

The 53-bp fragment containing the T7 promoter and target-specific spacer was synthesised as a double-stranded oligonucleotide (Metabion AG, Germany). The 80-bp crRNA/tracrRNA scaffold fragment and reagents for PCR and IVT were supplied by the Precision gRNA Synthesis Kit (Thermo Fisher Scientific, A29377). Fragments were assembled by PCR to generate 118-bp templates for IVT (Fig. 6).

Table 2. Modular design of gRNA DNA templates.

Element	Length (nt)	Sequence
T7 promoter sequence	18 nt	TAATACGACTCACTATAG
Target-specific spacer sequence	20 nt	NNNNNNNNNNNNNNNNNNNN
crRNA/tracrRNA constant region (underlined part priming with guide oligo)	80 nt	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCTTTT

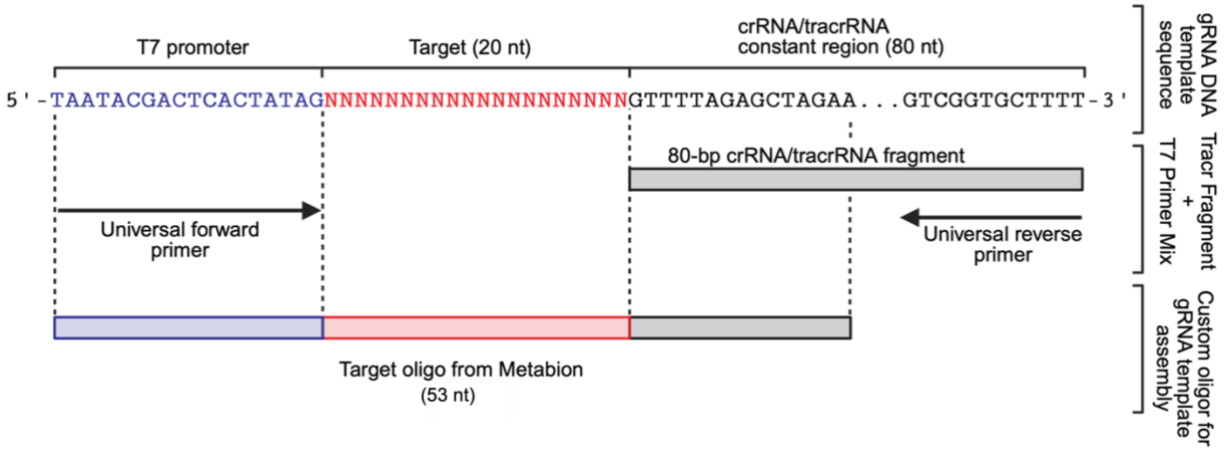


Figure 6. Modular architecture of gRNA DNA templates and PCR assembly strategy. *Each gRNA DNA template was designed as a modular construct consisting of three elements: (i) a T7 promoter sequence (18 nt, blue), required for transcription initiation by T7 polymerase, (ii) a target-specific spacer sequence (20 nt, red), corresponding to the genomic locus immediately upstream of a PAM motif, and (iii) the constant crRNA/tracrRNA scaffold (80 nt, black), which provides the structural domain necessary for Cas9 recognition and activity. Templates were assembled by PCR using a forward primer annealing within the T7 promoter region and a universal reverse primer complementary to the 3' end of the scaffold sequence.*

2.2.3 PCR assembly of gRNA DNA templates

Full-length gRNA DNA templates (118 bp) were generated by assembly PCR with the Precision gRNA Synthesis Kit (Thermo Fisher Scientific, A29377). Rather than the two overlapping oligonucleotides (Target F1/R1) recommended by the manufacturer, a single full-length target-specific oligonucleotide (0.3 µM final concentration, Metabion AG) was used to simplify the workflow and reduce primer-dimer formation (Fig. 6).

Each 25 µl PCR reaction contained the components listed in Table 3, and amplification was carried out under the cycling conditions summarised in Table 4.

Table 3. PCR assembly reaction composition.

Component	Amount	Final conc.	Source
Phusion™ High-Fidelity PCR Master Mix (2X)	12.5 µl	1X	Thermo Fisher
Tracr Fragment + T7 Primer Mix	1 µl	-	Thermo Fisher
Target-specific oligonucleotide	1 µl	0.3 µM	Metabion AG
Nuclease-free water	10.5 µl	-	Life Technologies

Table 4. gRNA DNA template assembly PCR cycling programme.

Cycle step	Temperature	Time	Cycles
Initial denaturation	98°C	10 seconds	1X
Denaturation	98°C	5 seconds	32X
Annealing	55°C	15 seconds	32X
Final extension	72°C	1 minute	1X
Hold	4°C	Hold	1X

Amplicons of the expected size (118 bp) were confirmed by 2 % agarose gel electrophoresis.

2.2.4 In vitro transcription of sgRNAs

Single guide RNAs (sgRNAs) were synthesised by *in vitro* transcription (IVT) from DNA templates using the Precision gRNA Synthesis Kit (Thermo Fisher Scientific, A29377). Each 20 µl IVT reaction was prepared according to the manufacturer's instructions, using the components listed in Table 5, and incubated at 37 °C for 3 h. Residual DNA was removed by adding 1 µl of RNase-free DNase I and incubating at 37 °C for 15 min. sgRNAs were purified using GeneJET™ RNA Purification Micro Columns provided by the kit. RNA integrity and concentration were assessed on an Agilent TapeStation 4150 using RNA ScreenTape and High Sensitivity RNA ScreenTape to confirm suitability for ribonucleoprotein (RNP) assembly.

Table 5. IVT reaction composition.

Component	Amount
NTP mix (25 mM each of ATP, GTP, CTP, UTP in Tris buffer)	8 µl
6 µL of unquantified gRNA DNA template from PCR product according to the kit protocol	6 µl
5X TranscriptAid™ Reaction Buffer	4 µl
TranscriptAid™ Enzyme Mix	2 µl

2.3 Ribonucleoprotein (RNP) assembly and electroporation

2.3.1 Preparation of Cas9-sgRNA complexes

Purified sgRNAs were complexed with *Streptococcus pyogenes* Cas9 (TrueCut™ Cas9 Protein v2, Thermo Fisher Scientific, A36498) in Buffer R (Neon™ Transfection System 10 µl Kit, Thermo Fisher, MPK1096). For each promoter, two sgRNAs (Table 6) were pooled at 0.75 µM each (1.5 µM total) and incubated with 0.48 µM Cas9 (Cas9:sgRNA molar ratio of about 0.32:1) for 20 minutes at room temperature. The total volume was 10 µl (composition in Table 6).

Table 6. Composition of Cas9-sgRNA RNP complexes (10 µl reaction)

Component	Amount	Final concentration in 10 µl	Notes
Cas9 (TrueCut™ Cas9 Protein v2)	4.8 pmol	0.48 µM	5 µg/µl stock
sgRNA (guide A)	7.5 pmol	0.75 µM	IVT product
sgRNA (guide B)	7.5 pmol	0.75 µM	IVT product
Total sgRNA	15 pmol	1.5 µM	Two-guide pool
Buffer R	to 10 µl	-	Neon kit buffer

2.3.2 Cell preparation and electroporation

H9 SBact cells were cultured to approximately 60% confluence, washed with PBS, and dissociated with Versene (0.48 mM EDTA in PBS, Gibco, 15040-066). Cells were pelleted at 300 x *g* for 5 minutes and resuspended in Buffer R containing 7.2 µM Alt-R™ Cas9 Electroporation Enhancer (Integrated DNA Technologies, 1075916, 200 µM in 1 x TE buffer). Cell suspensions were adjusted to 2.0 × 10^7 cells/mL and counted using a Cellometer K2 (Nexcelom).

For each electroporation, 5 µl of RNP was mixed with 5 µl of cell suspension (1.0 × 10^5 cells) and electroporated using the Neon™ Transfection System (10 µl tips, 1,500 V, 20 ms, 2 pulses, MAN0001447). Each sgRNA pool was delivered in duplicate (2 × 1.0 × 10^5 cells), and the electroporated cells were combined into a single Laminin-521-coated well in Essential 8 medium supplemented with 10 µM Y-27632. After 24 h, the ROCK inhibitor was removed, and cells were maintained in E8 with daily medium changes. Cells were expanded to 60-70 % confluence, corresponding to colony

diameters of 0.5-1.0 mm. At this point, the pools were frozen, tested, and used to initiate subcloning.

2.4 Clonal isolation of genome-edited hESC cells

For subcloning, pooled cultures were rinsed with 1 mL of PBS and incubated with 0.7 mL of Versene at 37 °C for 7 min until colony edges loosened. Colonies were gently detached, collected in 4 mL of Essential 8 medium, pelleted (300 x g for 5 min), and resuspended in 1 mL of E8. Cell viability and counts were determined using a Cellometer K2 (Nexcelom Bioscience). Cells were plated at 300 cells per 6-cm culture dish (≈ 10 cells/cm²) in 5 mL of E8 with 10 μM Y-27632. For each condition, two dishes with slightly different cell densities were seeded. After 24 hours, the ROCK inhibitor was removed, and cells were cultured for 6-8 days with daily E8 medium changes. Colonies measuring 0.5-1.0 mm with compact morphology, smooth edges, and phase-bright centres were selected for clonal picking.

For clonal expansion, 24-well plates were preloaded with 500 μl of E8 containing 10 μM Y-27632 per well. Individual colonies were aspirated in 20-30 μl of medium using a 200-μl tip and transferred to separate wells. Each colony was assigned a unique clone identifier. After overnight incubation, colonies were examined microscopically. Colonies that successfully reattached and began to spread were considered viable. Medium was replaced with 500 μl of E8 without Y-27632 and changed daily.

Over the next 4-5 days, colonies were gently passaged within the same plate using Versene. By days 8-10 after picking, colonies typically reached 70-90% confluence. At this point, cultures were split: a fraction of cells was harvested for genomic DNA extraction, while the remaining cells were expanded as clonal lines in fresh 24-well plates.

2.5 Genomic DNA extraction and TFAP2C promoter knockout validation

2.5.1 Genomic DNA extraction

Genomic DNA (gDNA) was isolated from individual clonal cultures in 24-well plates at 70-90% confluence. Wells were rinsed once with PBS and lysed in situ using the NucleoSpin Tissue kit (Macherey-Nagel, 740952.250). For each well, 200 μl of Buffer

T1 was added, and the lysate was transferred to 1.5 mL microcentrifuge tubes. Then, 25 µl of Proteinase K was added, and samples were incubated at 70 °C for 15 minutes.

DNA purification was performed according to the manufacturer's instructions, and DNA was eluted in 50 µl of buffer BE provided in the kit. DNA quality and purity were verified spectrophotometrically (NanoDrop). Acceptance criteria were A260/280 ratios of 1.8-2.0 and A260/230 ratios >1.8. A minimum of 90 ng of gDNA was allocated to each PCR reaction, with the remaining DNA stored at -20 °C.

2.5.2 Knockout validation primer design

Primers were designed relative to the CRISPR-Cas9 guide sites described in Section 2.2.1 (gRNA target selection) to detect engineered deletions within the *TFAP2C* promoter regions (Fig. 7). For each promoter, a pair of screening primers was positioned to amplify across the targeted interval (PCR1), enabling detection of deletion alleles by a shift in amplicon size relative to the wild-type product.

Nested PCR assays were used to confirm candidate clones and to generate high-quality junction-spanning fragments suitable for Sanger sequencing. For Promoter 2 (P2, Fig. 7, panel A), the first nested reaction reused the screening assay primer pair (PCR1) to preserve detection sensitivity, whereas the second nested reaction (PCR2) used internal primers to verify the deletion junction with higher specificity. For Promoter 1 (P1, Fig. 7, panel B), the nested workflow also reused the screening primers, but in an alternative forward-reverse configuration for the first nested reaction (PCR2), followed by a second nested reaction (PCR3) using fully internal primers to produce a compact, sequencing-ready amplicon.

Primer design was performed using Primer-BLAST (NCBI) with the GRCh38 human reference genome as the template. Design parameters specified melting temperatures of 60-62 °C, and *in silico* analysis predicted a single amplicon from wild-type H9 genomic DNA. All primer sequences are summarised in Table 7, and their genomic positions relative to exon 1 and the CRISPR-Cas9 cut sizes are illustrated in Figure 7. Oligonucleotides (Metabion AG) were supplied at a 100 µM stock concentration, diluted to 10 µM working stocks upon arrival, and stored at -20 °C until use.

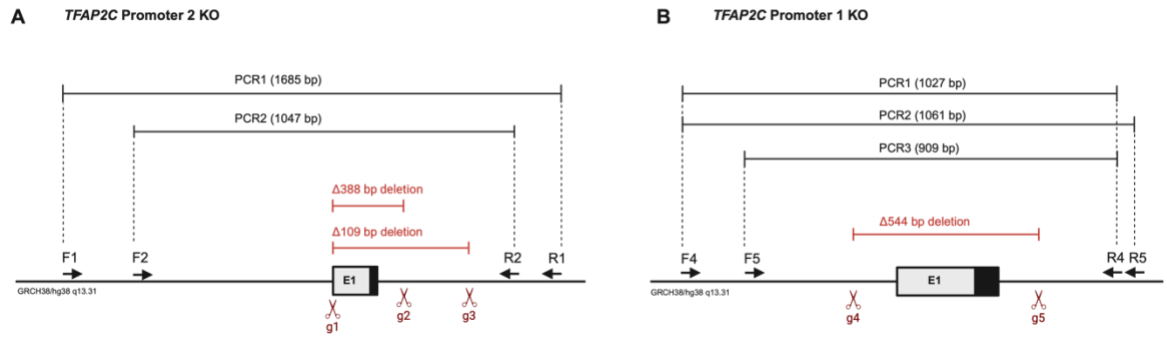


Figure 7. Primer placement and validation workflow for TFAP2C promoter knockouts. (A) *Promoter 2: screening (PCR1) and nested validation (PCR2) assays spanning CRISPR cut sizes g1-g3.* ***(B)*** *Promoter 1: screening (PCR1) and sequential nested reactions (PCR2-PCR3) confirming the Δ544 bp deletion between g4-g5.*

Table 7. Primer sets for TFAP2C promoter knockout validation.

Target	Assay	Primer ID	Sequence (5' -> 3')
Promoter 2	Screening, Nested PCR1	F1	GAAGCTGTTTGGGTTGCCTG
Promoter 2	Screening, Nested PCR1	R1	GCCGCATGCTCTTCGTTTAG
Promoter 2	Nested PCR2	F2	CCCGGAAACCTGTGATGTGA
Promoter 2	Nested PCR2	R2	CTGAAAGCCACAAAGGTGCC
Promoter 1	Screening (PCR1)	F4	CCTTAAGCCTCAGTGGGGTG
Promoter 1	Screening (PCR1)	R4	TTAGGAAAGCGCCCGGTAAG
Promoter 1	Nested PCR2	F4	CCTTAAGCCTCAGTGGGGTG
Promoter 1	Nested PCR2	R5	TTGGAGGAACGAGTGAAGGC
Promoter 1	Nested PCR3	F5	TTCCGAGAAAACGGGAAGCA
Promoter 1	Nested PCR3	R4	TTAGGAAAGCGCCCGGTAAG

Predicted wild-type and deletion product sizes are listed in Table 8.

Table 8. Expected amplicons for TFAP2C promoter deletions.

Target	Assay	Wild-type amplicon	Predicted knockout amplicon(s)	Deletion interval	gRNA pair(s)
P2	Screening, Nested PCR1	1685 bp	1576 bp (Δ109, g1-g2), 1297 bp (Δ388, g1-g3)	Δ109/Δ388	g1-g2/g1-g3
P2	Nested PCR 2	1047 bp	938 bp (Δ109, g1-g2), 659 bp (Δ388, g1-g3)	Δ109/Δ388	g1-g2/g1-g3
P1	Screening (PCR1)	1027 bp	483 bp	Δ544	g4-g5
P1	Nested PCR2	1061 bp	517 bp	Δ544	g4-g5
P1	Nested PCR3	909 bp	365 bp	Δ544	g4-g5

2.5.3 PCR reaction setup and cycling conditions

Genotyping reactions were optimised separately for validation of knockouts at Promoter 1 (P1) and Promoter 2 (P2). All assays used Phusion HF DNA Polymerase (Thermo Fisher Scientific) with distinct cycling programmes optimised for the corresponding amplicon lengths and primer configurations.

Promoter 1 (P1). Knockout validation for P1 was performed using a touchdown PCR protocol to enhance specificity and reduce nonspecific amplification. Screening (PCR1) and both nested reactions (PCR2-PCR3) used the same reaction mixture and cycling programme, with 1 µl of the PCR2 product serving as the template for the nested PCR3 reaction.

Table 9. TFAP2C P1 KO genotyping PCR (touchdown)

Component	Manufacturer	Volume (50 µl reaction mix)
5x Phusion HF buffer	Thermo Fisher Scientific	10.0 µl
2.5 mM dNTP mix	BioNordika	4.0 µl
10 µM forward primer	MetaBion AG	2.5 µl
10 µM reverse primer	MetaBion AG	2.5 µl
Phusion HF DNA polymerase (2 U/µl)	Thermo Fisher Scientific	0.5 µl
Genomic DNA (90 ng) (PCR1/PCR2) /1 µl template from the previous PCR2 reaction (PCR3)	-	1-3 µl
Nuclease-free water	Life Technologies	to 50 µl

Table 10. P1 genotyping touchdown PCR cycling programme.

Step	Temp (°C)	Time
1 Initial denaturation	98	3 min
2 Denaturation	98	10 s
3 Annealing	72, decrease 0.1 °C per cycle	30 s
4 Extension	72	1 min
Repeat steps 2-4 x 29		
5 Denaturation	98	10 s
6 Annealing	69	30 s
7 Extension	72	1 min
Repeat steps 5-7 x 9		
8 Final extension	72	8 min
9 Hold	4	∞

Promoter 2 (P2). Knockout validation for P2 used a two-stage nested PCR approach with a fixed annealing temperature (Table 11 and 12). The same reaction composition was used for all assays, with the PCR1 product serving as template for PCR2 (Nested-2).

Table 11. TFAP2C P2-KO genotyping PCR.

Component	Manufacturer	Volume (20 µl reaction mix)
5x Phusion HF buffer	Thermo Fisher Scientific	4.0 µl
2.5 mM dNTP mix	BioNordika	1.6 µl
10 µM forward primer	MetaBion	1.0 µl
10 µM reverse primer	MetaBion	1.0 µl
Phusion HF DNA polymerase (2 U/µl)	Thermo Fisher Scientific	0.2 µl
Genomic DNA (90 ng) (PCR1) /1 µl template from the previous PCR1 reaction (PCR2)		1.0 µl
Nuclease-free water	Life Technologies	to 20 µl

Table 12. P2 genotyping PCR cycling programme.

Step	Temp (°C)	Time
1 Initial denaturation	98	30 s
2 Denaturation	98	10 s
3 Annealing	63	30 s
4 Extension	72	1 min
Repeat steps 2-4 x 29		
5 Final extension	72	8 min
6 Hold	4	∞

2.5.4 Sanger sequencing

PCR products were resolved on 1% agarose in 0.5x TBE, and knockout-diagnostic bands of the expected nested-PCR sizes were excised and purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel), then eluted in 15 µl. DNA concentration and purity were verified spectrophotometrically.

Purified amplicons were submitted to EuroFins Genomics for Sanger sequencing under standard template/primer conditions (15 ng/µl template, 5 µM primer). Sequencing used the inner forward primers from the nested reactions (F5 for Promoter 1, F2 for Promoter 2). The corresponding inner reverse primers were included as required to confirm zygosity. Chromatograms (.ab1) were quality-checked and aligned to the TFAP2C locus (GRCh38) using ApE and Benchling. Biallelic knockouts were confirmed when only the junction product was detected, and the wild-type signal was absent. Mixed traces indicative of mosaicism or heterozygosity prompted re-sequencing and, when necessary, re-cloning.

2.6 Pluripotency qualification

Following CRISPR-Cas9 editing and clonal isolation, biallelic deletion lines targeting either the P1- or P2-associated region were assessed for maintenance of pluripotency using combined morphological and functional criteria.

2.6.1 Morphological assessment

Pluripotency pre-qualification was first confirmed by examining colony morphology under undifferentiated culture conditions. Morphologically pluripotent colonies were required to exhibit a high nuclear-to-cytoplasmic ratio, prominent nucleoli, and compact colony borders with sharp, undifferentiated edges, lacking peripheral flattening or spontaneous differentiation. Cells were maintained on Laminin-521 in Essential 8 medium and passaged using Versene/EDTA (0.48 mM) to ensure gentle dissociation and preservation of undifferentiated state as described in section 2.1.2.

2.6.2 Functional trilineage assessment

Functional pluripotency was then validated by direct trilineage differentiation, followed by lineage-specific immunocytochemistry (ICC). Each lineage differentiation protocol followed the standardised workflow summarised in Figure 8 and Table 13, which details coating preparation, media composition, timing, small-molecule treatments, passaging steps, and endpoint markers. All steps were performed exactly as listed. Direct ICC for pluripotency markers in the undifferentiated state was not performed.

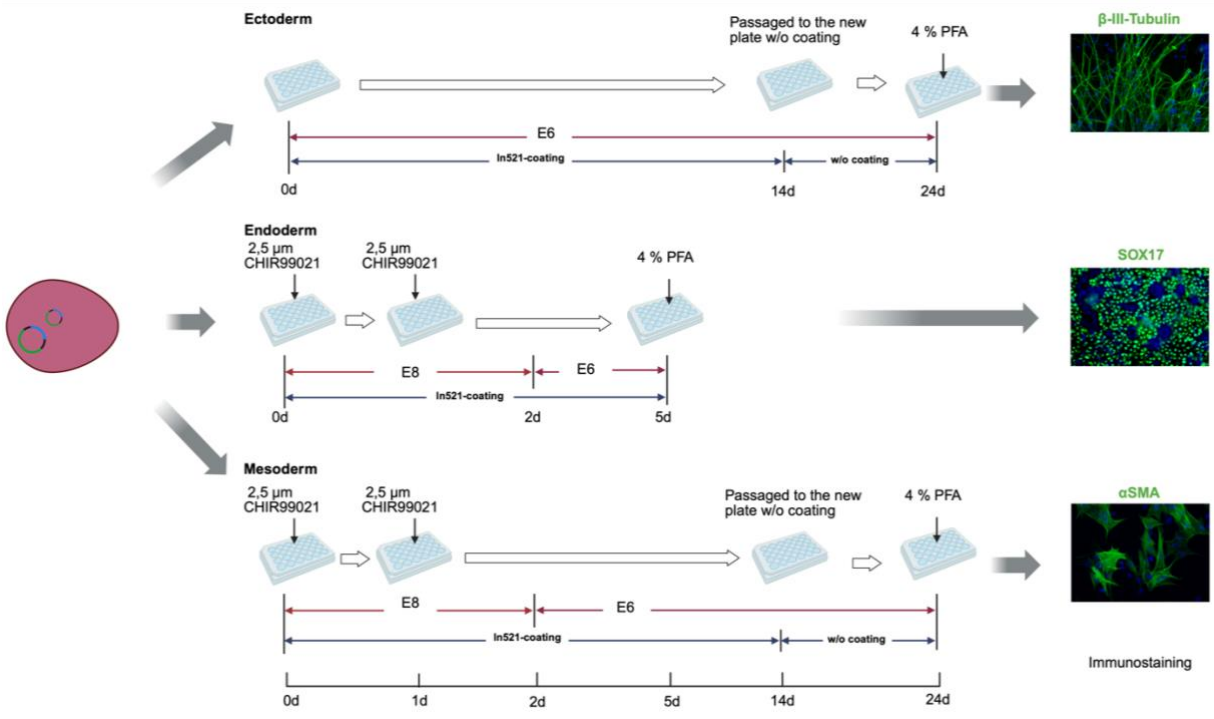


Figure 8. Functional trilineage differentiation workflow and endpoints.

Ectoderm differentiation was induced by switching to E6 medium. Cultures were maintained on Ln521 in E6 until day 14, then gently dissociated with TrypLE Express (Life Technologies, Ref#12604021) and seeded onto uncoated plates in E6 to promote neural maturation. On day 24, terminal ectoderm identity was supported by cell morphology and β-III-tubulin staining.

Endoderm differentiation was induced by a 48-h exposure to the GSK3β antagonist CHIR99021 (2.5 μM) in E8, followed by 3 days in E6. On day 5, definitive endoderm was indicated by nuclear SOX17 expression.

Mesoderm differentiation was similarly induced by a 48-h exposure to the GSK3β antagonist CHIR99021 in E8, followed by 12 days in E6 on LN521, then transfer to uncoated plates and a further 10 days in E6. By day 24, α-smooth muscle actin (α-SMA) staining revealed organised filamentous networks typical of mature mesoderm.

Table 13. Functional trilineage differentiation workflow and endpoints.

Component	Coating	Medium and schedule	Passaging step	Total duration	ICC marker
Ectoderm	LN521, 10 µg/mL in PBS, 0.250 mL per 24-well plate, 2 h at 37 °C PBS with Mg2+/Ca2+	E6 on LN521 (days 0-14), E6 on uncoated plates (days 14-24)	Day 14, transferred to uncoated plates	24 days	β-III-tubulin
Endoderm	LN521 as above	E8 + CHIR99021 (day 0-2), switch to E6 (days 2-5)	None	5 days	SOX17
Mesoderm	LN521 as above	E8 + CHIR99021 (day 0-2), switch to E6 (days 2-14), E6 on uncoated plates (days 14-24)	Day 14, transferred to uncoated plates	24 days	α-SMA

Cells were fixed with 4 % paraformaldehyde (PFA) for 30 minutes at room temperature. Permeabilisation was performed using 0.2 % Triton X-100 for 15 minutes, followed by three washes with PBS. Samples were then blocked with 4 % fetal bovine serum (FBS) and 0.1 % Tween-20 for 1 hour at room temperature. Primary antibodies (listed below) were diluted in freshly prepared blocking buffer and incubated at 4 °C for 48 hours.

After primary antibody incubation, cells were washed three times with PBS and incubated with the appropriate secondary antibodies together with 20 µM Hoechst 33342 (Thermo Fisher Scientific, 62249) for 1 h at room temperature in the dark. Hoechst was used to visualise cell nuclei. Following secondary antibody incubation, cells were washed three times with PBS and imaged using a Nikon Eclipse Ts2 fluorescence microscope.

Table 14. Primary antibodies for trilineage differentiation.

Gene	Host species	Dilution	Manufacturer	Product code
β-III-tubulin (TUBB3)	mouse	1:1000	Invitrogen	14-4510-82
SOX17	rabbit	1:400	Cell Signaling	81778
α-SMA	mouse	1:1000	Merck / Sigma	A2547-100UL

Table 15. Secondary antibodies for trilineage differentiation.

Antibody	Dilution	Manufacturer	Product code
Donkey anti-Rabbit IgG (H+L) Alexa Fluor 488	1:500	Invitrogen	A21206
Donkey anti-Mouse IgG (H+L) Alexa Fluor 488	1:500	Life Technologies	A21202

2.7 Promoter-specific activation of TFAP2C using CRISPRa and induction of trophoblast factor-expressing cells

2.7.1 Design of TFAP2C CRISPRa guide RNAs

Promoter-selective CRISPR activation (CRISPRa) was used to upregulate either the *TFAP2C* promoter 1 or promoter 2 in promoter wild-type and promoter-knockout H9 SBact hESCs. Three gRNAs were designed for each promoter using the Benchling design platform. Candidate guides were selected based on predicted on-target efficiency and proximity to the intended promoter region. Activation guides targeting P1 were positioned upstream of the canonical transcription start site (TSS), whereas P2-specific activation guides were designed to activate the novel promoter P2. All activation guide sequences and their strand orientations are provided in Table 16.

Table 16. CRISPR activation guide sequences

CRISPR activation guide	Sequence	Strand
TFAP2C P1 g1	ACGCTGTGAGGGACGAAGCG	-
TFAP2C P1 g2	AATCTCAGGTGTGTTTACGG	+
TFAP2C P1 g3	GGGGAGCGACTCAACAACCG	-
TFAP2C P2 g1	CAGAGTTTATAGTTTCCAAG	+
TFAP2C P2 g2	ATCCCTTGCCAGTAAGAGGT	+
TFAP2C P2 g3	GTTCAGCTGAAATTGAACCA	-

The promoter 2-derived transcript contains a distinct exon 1 that splices directly to exon 2, whereas the canonical transcript from promoter 1 initiates at the consensus exon 1, which also splices to exon 2. Because these first exons are promoter-specific and mutually exclusive, deleting P1, including its exon 1, does not disrupt production of the P2 transcript, as the P2 does not utilise any sequence from P1 or the canonical exon 1. Conversely, deleting P2-associated exon 1 does not interfere with generation of the P1-derived transcript, which relies solely on the canonical promoter and its exon 1. In both cases, promoter-specific deletions preserve the structural integrity of the alternative transcript, although transcriptional output may still change if removal of either promoter alters local chromatin accessibility or modifies enhancer-promoter interactions.

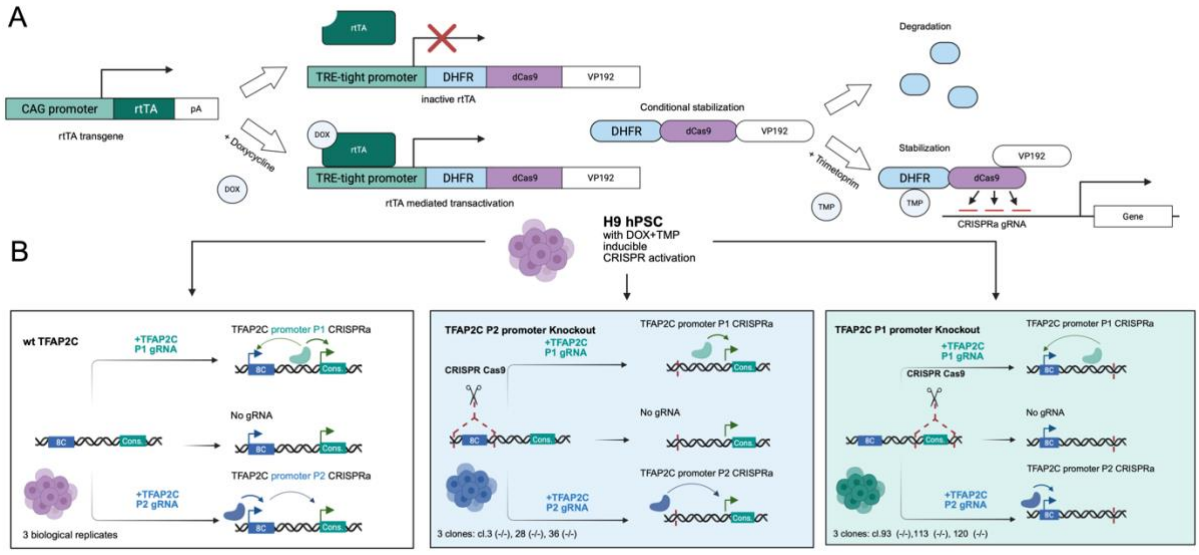


Figure 9. Schematic image illustrates the inducible CRISPRa system used in this study*, including rtTA-dependent TRE promoter activation, DHFR-mediated conditional stabilisation with trimethoprim, and the dCas9-VP192 transcriptional activation module. Illustration created by the author.*

Figure 9 shows the inducible CRISPRa system used in this study. The schematic illustrates the TRE-tight promoter activated by rtTA in the presence of doxycycline, the DHFR destabilisation domain stabilised by trimethoprim, and the DDdCas9-VP192 activation module, which together enable small-molecule control of CRISPRa activity.

All CRISPRa target sites were deliberately positioned upstream of the genetic regions deleted during CRISPR/Cas9-mediated P1 and P2 promoter deletions. Consequently, every CRISPRa target site remains fully intact and accessible to dCas9-VP192 in wild-type, P1-KO, and P2-KO lines. Because none of the activation guide sites are deleted during promoter removal, promoter-selective CRISPRa and cross-activation remain possible in all genetic backgrounds, depending on the local chromatin context.

2.7.2 CRISPRa vector construction

All CRISPRa guide sets were cloned into PiggyBac (PB) gRNA expression vector containing a puromycin resistance cassette. Cloning was performed using Golden Gate assembly, with BsaI-digestion and ligation of modular gRNA transcriptional units. The assembled constructs were verified by restriction digest analysis and Sanger sequencing (Eurofins Genomics, Germany).

2.7.3 Establishment of CRISPRa cell lines

The H9 SBact cl.11 hESC line, which stably expresses DDdCas9-VP192 under a TetON promoter together with CAG-rtTA, served as the CRISPRa-competent parental line. Promoter-specific knockout lines (P1-KO and P2-KO) were generated on this background to ensure that all cells retained full inducible CRISPRa capacity. After the knockout lines were established, the TFAP2C P1- and P2-targeting gRNA PiggyBac constructs were introduced into both KO lines by electroporation. The corresponding wild-type CRISPRa lines carrying the same guide sets had been produced earlier and were used as reference controls, whereas all CRISPRa-competent KO lines were produced as part of this study.

To prepare cells for electroporation, cultures were dissociated with Versene (Thermo Fisher Scientific). After dissociation, 1 x 10^6 cells were counted using the Cellometer Auto T4 cell counter (Nexcelom Bioscience), centrifuged at 300 x g for 5 min, and resuspended in 110 µl of Neon R-buffer. The cell suspension was then mixed with the plasmid solution. The PiggyBac transposition plasmid mix contained 1 µg PB-GG-

TFAP2C-P1-3g-PP or PB-GG-TFAP2C-P2-3g-PP together with 0.5 µg of the CAG-hyPBase-IRES-Venus transposase plasmid. Electroporation was performed using the Neon Transfection System (Invitrogen, MPK10096) with a 100 µl electroporation tip, according to the manufacturer's instructions. The optimised electroporation parameters were 1100 V, 20 ms, and two pulses.

Following electroporation, cells were plated onto Laminin-521-coated 6-well plates in E8 medium supplemented with 10 µM Y-27632 (Cayman, 10005583) and allowed to recover for 24 h. Medium was then replaced with fresh E8 without Y-27632, and YFP expression from the transposase plasmid was imaged to confirm successful transfection.

2.7.4 Antibiotic selection of CRISPRa guides

Stable integrants were enriched with puromycin (0.5 µg/ml, Thermo Scientific, J67236.XF) for five days, then cultured in antibiotic-free E8 for 2 days before passaging and cryopreservation. Plasmid maps for all PiggyBac constructs, including PB-GG-TFAP2C-P1-3g-PP, PB-GG-TFAP2C-P2-3g-PP and CAG hyPBase-IRES-Venus, are presented in Appendix 1, Supplementary Figures 1-5.

2.7.5 Induction of TFAP2C expression

For CRISPRa, cultures were treated with 2 µg/ml doxycycline (Fisher Bioreagents, BP2653-1) and 1 µM trimethoprim (Fisher Scientific, 15345602) for 48 h. During and after induction, cells were cultured in Essential 6 medium (E6, Gibco, Thermo Fisher Scientific, A1516401), providing a defined baseline environment to assess transcription factor-driven trophoblast priming without exogenous signalling. Other culture conditions were identical to those used for undifferentiated H9 cells.

2.8 Immunocytochemistry

Cells were fixed in 4 % paraformaldehyde for 30 min, washed in PBS, and permeabilised with 0.2 % Triton X-100 (Sigma-Aldrich, T9284) in PBS for 15 min at RT. After three PBS washes, samples were blocked for 1 h at RT in 4 % FBS and 0.1 % Tween-20 (Sigma-Aldrich, P1379) in PBS. Primary antibodies (Table 17) were diluted in blocking buffer and incubated for 48 h at 4 °C. After three washes, cells were incubated for 1 h at RT with the appropriate secondary antibodies together with 20 µM

Hoechst 33342 (Thermo Fisher Scientific, 62249), washed three times with PBS, and imaged using a Nikon Eclipse Ts2 fluorescence microscope.

Table 17. Primary Antibodies

Gene	Host species	Dilution	Manufacturer	Product code
GATA2	Rabbit	1:400	Sigma	HPA005633
hCGB	Rabbit	1:400	Invitrogen	PA5-58598
HLA-G	Mouse	1:400	Santa Cruz	sc-21799
KRT7	Mouse	1:800	Thermo Fisher	MA1-06316
TFAP2C	Goat	1:1000	R&D Systems	AF5059

Table 18. Secondary Antibodies

Antibody	Dilution	Manufacturer	Product code
Donkey anti-Goat IgG (H+L) Alexa Fluor 488	1:500	Invitrogen	A11055
Donkey anti-Rabbit IgG (H+L) Alexa Fluor 488	1:500	Invitrogen	A21206
Donkey anti-Mouse IgG (H+L) Alexa Fluor 488	1:500	Life Technologies	A21202
Donkey anti-Rabbit IgG (H+L) Alexa Fluor 594	1:500	Invitrogen	A21207
Donkey anti-Mouse IgG (H+L) Alexa Fluor 594	1:500	Invitrogen	A21203

2.9 Imaging and image processing of protein expression

All brightfield and fluorescence images were acquired using a Nikon Eclipse TS2 microscope under fixed exposure, illumination and gain settings. Fluorescence channels were merged in Fiji (Schindelin et al., 2012), and only linear brightness and contrast adjustments were applied uniformly across all images.

2.10 Quantitative reverse transcription PCR (RT-qPCR)

Cells were lysed directly on culture plates, and total RNA was extracted using the NucleoSpin RNA Plus Kit (Macherey-Nagel, 740984). RNA concentrations were measured using a DS-11 FX+ spectrophotometer (DeNovix).

Reverse transcription was performed using 200 ng of RNA. RNA was diluted in 11.3 µl of nuclease-free water (Invitrogen, AM9937) and denatured for 1 min at 65 °C. The final 20 µl RT reaction contained 4 µl of RT buffer (Promega, M531A), 2.5 µl of 2.5 mM dNTP mix (BioNordika, BN-1006-08), 1 µl of Oligo(dT) primer (Thermo Scientific, SO132), 0.2 µl of random hexamers (Promega, C118A), 0.5 µl of RNase inhibitor (Thermo Scientific, EO0381), and 0.5 µl of M-MLV reverse transcriptase (Promega,

M170A). Reactions were incubated for 90 min at 37 °C and then inactivated for 5 min at 95 °C. cDNA was diluted 1:5 prior to RT-qPCR.

RT-qPCR reactions (10 µl) contained 4 µl cDNA, 2 µl 5x buffer (7.5 mM, Thermo Scientific, F524), 0.2 µl 10 mM dNTPs, 0.5 µl 20x EvaGreen dye (Biotium, 31000), 0.5 µl each of forward and reverse primers (10 µM, Table 19), 0.2 µl Phire Hot Start II DNA polymerase (Thermo Scientific, F122L), and nuclease-free water. Primers were previously validated in the laboratory for RT-qPCR suitability and product specificity was confirmed by gel electrophoresis (Appendix 7). Each biological sample was run in technical duplicates on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) using the cycling programme shown in Table 20. Valid duplicate Cq values were averaged, whereas individual wells showing clear amplification failure were treated as missing. Four no-template controls (NTCs) with nuclease-free water were included for each tested gene, and no amplification was observed in NTCs.

Table 19. RT-qPCR primers

Gene	Forward primer (5'-> 3')	Reverse primer (5'-> 3')	Amplicon (bp)
KRT7	ACATCGAGATCGCCACCTAC	ACTGCCACCAGTGGAATTCAT	110
GATA2	CCACTACCTGTGCAATGCCT	AATTTGCACAACAGGTGCCG	119
HLA-G	TGTCCTTGCAGCTGTAGTCA	AAGGGACTTGCCACTCAGTC	141
hCGB	GCTGGACCAGTGAGAGGAGAG	CAAGGCCACCAGGAGGTGA	136
PPIG	TCTTGTCAATGGCCAACAGA	GCCCATCTAAATGAGGAGTT	84
TFAP2C P2 variant	GGCACCATGAGCTTCCAGTT	TATTCGGCGACTCCAGTGTG	140
TFAP2C P1 variant	AGTACGAAGAGGACTGCGAG	TATTCGGCGACTCCAGTGTG	133

Table 20. RT-qPCR programme.

Step	Temp (°C)	Time
Step 1: Initial denaturation	98 °C	30 s
Step 2: Denaturation	98 °C	5 s
Step 3: Combined annealing and extension	63 °C	8 s
Step 4: Plate read		
Repeat steps 2-4 x 40 times		

2.11 Statistical testing of RT-qPCR results

Relative transcript abundance was calculated using the 2^-ΔΔCt method (Livak and Schmittgen, 2001), with peptidylprolyl isomerase G (PPIG) as the endogenous reference gene, which showed no condition-dependent shifts in raw Cq values, and WT no-guide day 0 as the calibrator. Undetermined reactions were treated as missing data before ΔCt calculation. The plotted expression values represent mRNA abundance relative to WT no-guide day 0.

Primary statistical testing was performed using blocked linear models fitted to ΔCt values to account for biological variation between parental clones or WT passages (Krzywinski and Altman, 2014a). Models included biological block, guide condition, differentiation day as a categorical factor, and their guide-by-day interaction (Krzywinski and Altman, 2014b). Predefined contrasts compared P1 and P2 CRISPRa with the matched no-guide condition at each analysed day. Undetermined and failed reactions were treated as missing before modelling. P-values were two-sided and adjusted for multiple testing using the Benjamini-Hochberg false-discovery-rate procedure (Benjamini and Hochberg, 1995). Correction was applied separately to promoter-resolved *TFAP2C* assays and to the *GATA2* / *KRT7* marker assays. Statistical significance was indicated by adjusted q-values (*q < 0.05, **q < 0.01, ***q < 0.001). P1-transcript measurements in P1-KO cells and P2-transcript measurements in P2-KO cells were excluded from primary statistical testing because the corresponding region is deleted. All target/background/guide/time-point groups contained three biological replicates except for a single failed *GATA2* d5 replicate. Heteroscedasticity-robust HC3 tests were used as sensitivity analyses (Long and Ervin, 2000), and effects that weren't supported were interpreted cautiously.

Secondary promoter-dependence tests directly compared CRISPRa *KRT7* and *GATA2* responses between promoter-intact and promoter-deletion backgrounds. The baseline-adjusted log2 values (guide – no-guide) on the analysed day, minus (guide – no-guide) on day 0, were compared using two-sided Welch tests (Welch, 1947), followed by Benjamini-Hochberg correction, within the secondary promoter-dependence tests.

2.12 Figure 2 data processing

Figure 2 combines three data sources describing TFAP2C promoter-resolved transcription-far-end (TFE) usage and total gene expression during human preimplantation development. Transcript models in Figure 2A were taken from the Denis/Torre isoform-resolved preimplantation embryo dataset (GSE190548) and were plotted on GRCh38 using TALON transcripts TALONT000299008 (P2) and TALONT000299125 (P1). Quantitive promoter-resolved TFE percentages in Figure 2B were obtained from Töhönen et al. (2015) (PRJEB8994/ERP010048), and relevant hg19 STRT/TFE intervals were mapped to GRCh38 using UCSC LiftOver (Kent source tree v502). P2-positive cells corresponded to FE337643, FE337644, FE330330, FE337645 and FE337646, and P1-positive cells to FE337647 and FE337648. Total TFAP2C expression in Figure 2C was taken from Yan et al. (2013) (GSE36552), using gene-level RPKM values for NM_003222. Values were plotted as log10(RPKM + 1) without an expression threshold and were used only to show gene-level developmental timing.

2.13 Research ethics

2.13.1 Human embryonic stem cell research and genetic modification

The experimental work was conducted using the established H9 human embryonic stem cell line (WA09; WiCell Research Institute, Madison, WI, USA; RRID:CVCL_9773; hPSCreg WAe009-A). No embryos were used for stem cell derivation in this study, and no new human participants, patient samples or animal material were recruited or collected. Stem cell research involving the H9 cell line was conducted under approval from the Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS/2064/2019).

The ethical considerations associated with human embryonic stem-cell research were addressed by restricting the experimental work to an established, registered cell line and by conducting it under the applicable institutional approval. The study involved genetic modification of H9-derived cells using CRISPR/Cas9, CRISPR activation and transposon-based expression vectors. Genetic modification and the handling of the engineered hESC lines were carried out at Folkhälsan Research Center in accordance with the Finnish Gene Technology Act and the applicable institutional GMO notification and authorisation procedures. Work with viable genetically modified cells was performed using designated GMO class 1 laboratory procedures and approved laminar-flow cell-culture cabinets. GMO material and associated consumables were disposed of through the designated GMO-waste-

management system in accordance with Biomedicum 1 institutional procedures to minimise environmental exposure of GMO material.

The genome-editing experiments were conducted solely for research purposes to investigate promoter function. The engineered cells were maintained as in vitro experimental models and were not intended for reproductive, implantation or therapeutic use. Genomic integrity and genome-wide off-target effects were not comprehensively assessed, limiting conclusions regarding the broader genetic consequences of the engineered variants.

2.13.2 Data protection and research-data management

The experimental data generated in this study consisted of microscopy images, RT-qPCR measurements, genotyping PCR results and Sanger sequencing data from established hESC lines. These datasets did not contain any information or other direct personal identifiers. Experimental data were stored on a local computer and on an access-controlled University of Helsinki network drive. Data handling followed the University of Helsinki and Folkhälsan Research Center research data management practices.

2.13.3 AI use declaration

Grammarly 1.189.2.0 and OpenAI GPT-5.6 Sol were used during thesis preparation for grammar correction and proofreading. All AI-generated suggestions were independently evaluated by the author. Generative AI was not used to generate or modify experimental data, and the author remains responsible for all scientific analyses, references, and final wording.

3 Results

3.1 Generation and molecular validation of TFAP2C promoter-specific knockout hESC lines

3.1.1 Design and bulk performance of CRISPR/Cas9 editing

To examine promoter-specific control of *TFAP2C*, human embryonic stem cells (H9-SBact) were engineered to carry deletions of either the canonical promoter 1 (P1) and its exon 1, or the alternative exon 1 associated with the upstream promoter 2 (P2). The shared downstream coding exons were preserved, while the promoter-specific first exon and its corresponding N-terminal contribution were disrupted. The edits altered promoter-associated transcript structure while preserving the shared downstream *TFAP2C* coding region. Because each isoform contains a distinct, promoter-specific first exon, the isoform-specific deletion in each line is directly visible at the exon 1 level.

Dual-sgRNA ribonucleoprotein complexes (RNPs) were assembled from recombinant Cas9 protein and in vitro-synthesised sgRNAs, then electroporated into hESCs. Successful editing in bulk populations was first confirmed by PCR using primers flanking each target interval. For both promoter regions, edited pools showed clear shorter amplicons at the expected positions alongside the wild-type (WT) band, consistent with successful generation of deletion alleles. For P2, deletions of Δ109 (g1/g2) and Δ388 bp (g1/g3) were detected, confirming that both sgRNA combinations were active and capable of producing the intended excisions (Figure 11). For P1, the expected Δ544 bp deletion was observed (Figure 11). This PCR array detected no additional major amplicons.

The persistence of WT amplicons in edited pools indicated that RNP delivery did not fully saturate the population and that a proportion of alleles remained unedited. This justified clonal isolation followed by detailed genotyping to resolve allele-specific editing outcomes.

3.1.2 Clonal isolation and screening

Single cell derived hESC colonies were obtained by low-density seeding of edited pools and manual picking of colonies with undifferentiated morphology. Genomic DNA (gDNA) from each clone was subjected to diagnostic PCR spanning the targeted P1 or P2 intervals using the primer sets described in the Materials and Methods.

Table 21. Summary of genome-editing outcomes.

Target	Deletion	Biallelic deletions ± SE	heterozygous deletions ± SE	Viability
P1 (g4/g5)	544 bp	3/29 (10.3 % ± 5.7 %)	17/29 (58.6 % ± 9.1 %)	60.0 %
P2 (g1/g2)	109 bp	1/17 (5.8 % ± 5.7 %)	15/17 (88.2 % ± 7.8 %)	70.1 %
P2 (g1/g3)	388 bp	2/17 (11.8 % ± 7.8 %)	14/17 (82.4 % ± 9.3 %)	70.1 %

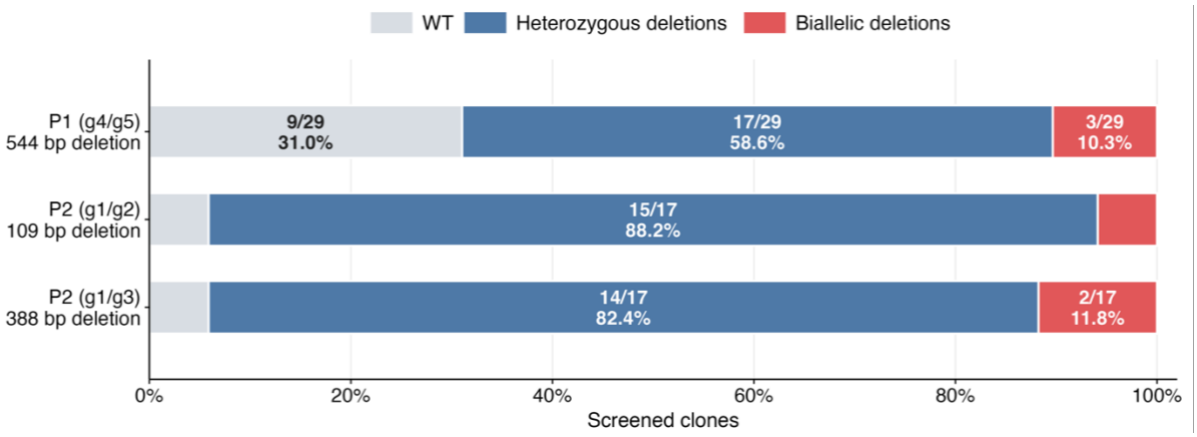


Figure 10. Summary of genome-editing outcomes. *P1 (g4/g5) deletion, n=29; P2 (g1/g2) deletion, n=17; and P2 (g1/g3) deletion, n=17.*

P1 editing (g4/g5) produced biallelic deletions in 10.3 % ± 5.7 % of clones and heterozygous deletions in 58.6 % ± 9.1 %. P2 editing was similarly efficient: the g1/g2 combination yielded 5.8 % ± 5.7 % biallelic deletions and 88.2 % ± 7.8 % heterozygous deletions, while g1/g3 generated 11.8 % ± 7.8 % biallelic deletions and 82.4 % ± 9.3 % heterozygous deletions.

These outcomes confirmed efficient sgRNA activity at both promoters and provided a pool of well-defined deletion lines for downstream functional studies.

3.1.3 Nested PCR and Sanger sequencing confirm targeted promoter-region deletions

Nested PCR with internal primers produced clean, sequencing-quality amplicons spanning the predicted junctions (Figure 11). P1 deletion alleles produced the expected 517 bp amplicon, corresponding to a 544 bp excision. P2 deletions yielded 1297 bp (Δ388 bp) and 1576 bp (Δ109 bp) amplicons, confirming targeted deletions of the expected approximate sizes with both sgRNA designs.

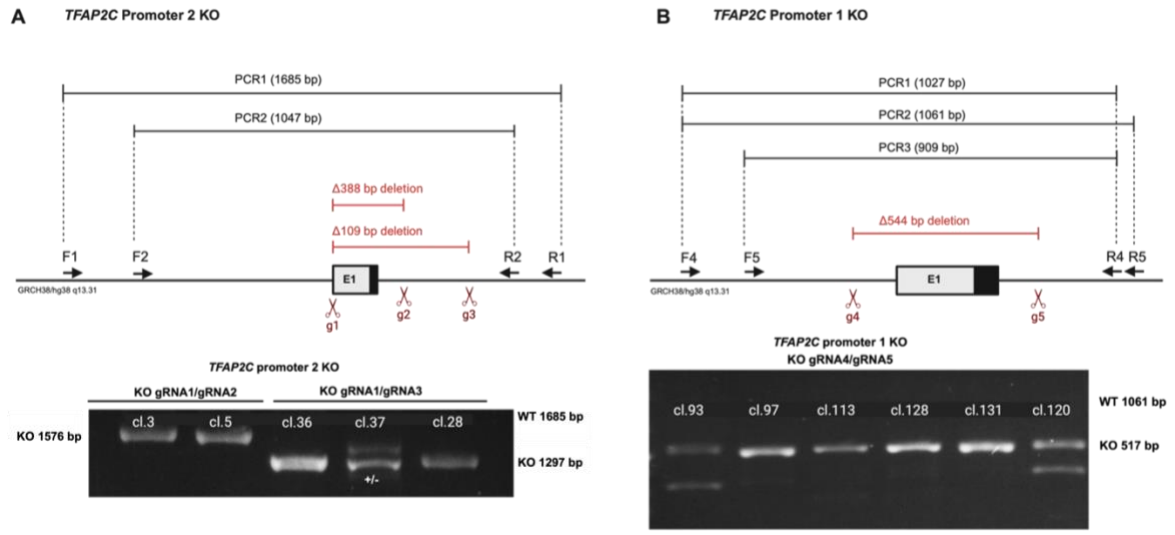


Figure 11. PCR-based validation of CRISPR/Cas9-mediated deletions in TFAP2C promoter/first exon regions. (A) *Schematic representation and genotyping of promoter 2 knockout using gRNA pairs g1/g2 and g1/g3. PCR amplification with primers F1-R1 produced wild-type (WT) and shorter knockout fragments of 1576 bp (Δ109 bp) and 1297 bp (Δ388), corresponding to targeted deletions within the targeted promoter region. Representative agarose gel electrophoresis confirms the expected amplicon size shifts for edited clones compared with the WT control.* ***(B)*** *Schematic representation and genotyping of TFAP2C promoter 1 KO using gRNA pair g4/g5. PCR with primers F4-R5 detected the expected 517 bp KO fragment corresponding to a 544 bp deletion relative to the 1061 bp WT product.*

Sanger sequencing confirmed targeted editing at both the *TFAP2C* P1 and P2 regions, with the observed junctions consistent with paired Cas9 cleavage followed by non-homologous end joining (Figure 12). In total, six independent TFAP2C knockout clones were analyzed, comprising three P2 clones (cl.3, cl.28, and cl.36) and three P1 clones (cl.93, cl.113 and cl.120). All six clones showed biallelic editing. In the P2 clones, deletions ranged from 100 to 405 bp. Clone 3 carried 109- and 113-bp deletions, clone 28 carried 391- and 405-bp deletions and clone 36 carried a 388-bp deletion on both alleles. Small 1-2 bp insertions were also detected in three of the six P2 alleles. In the P1 clones, the expected 544-bp deletion was present on at least one allele of each clone. Clone 113 carried the 544-bp deletion on both alleles, whereas clones 93 and 120 carried larger deletions of 738 bp and 681 bp, respectively, on their second alleles,

corresponding to additional deletions of 194 bp and 137 bp. These biallelic knockout clones were used for downstream functional assays to avoid confounding effects from residual wild-type *TFAP2C* transcripts.

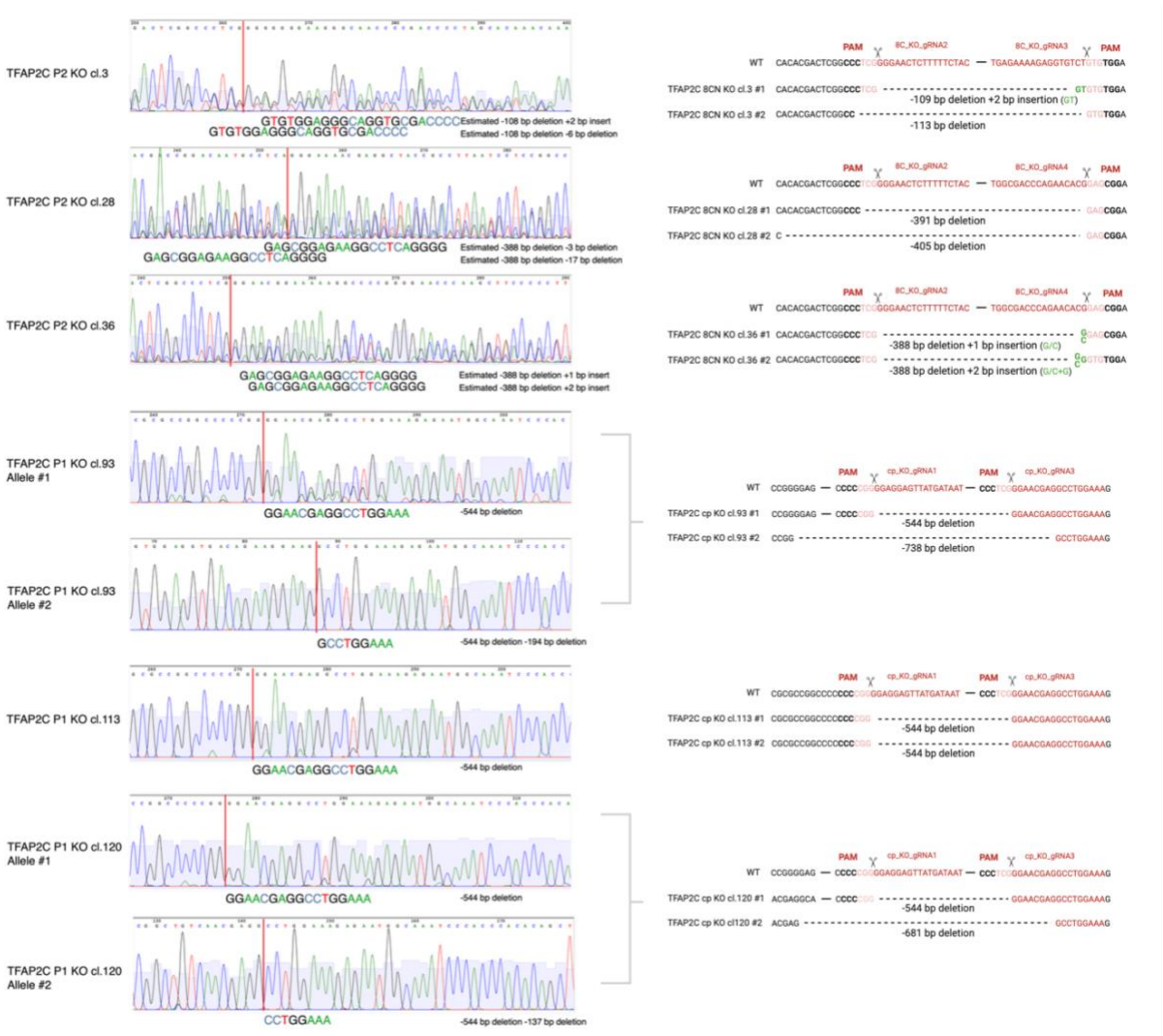


Figure 12. Sanger sequencing analysis of TFAP2C promoter knockout clones. *Representative chromatograms and corresponding sequence alignments illustrating CRISPR-Cas9-induced deletion variants in TFAP2C knockout clones. For each clone, Sanger traces (left) show disrupted signal and mixed peaks at the predicted Cas9 cut site, indicated by the vertical red line. Alignment panels (right) display the wild-type reference sequence alongside estimated detected alleles, highlighting deletions and insertions generated by non-homologous end joining. P2 KO lines exhibit compound editing outcomes, including deletions ranging from 100 bp to 400 bp and occasional small insertions. P1 KO clones show consistent biallelic editing, with recurrent 544 deletions and additional larger deletions depending on the allele. PAM motifs and gRNA target sequences are annotated, confirming editing at the intended locus.*

3.1.4 Pluripotency and trilineage potential of TFAP2C promoter knockout lines

All selected clones retained typical hESC morphology, forming compact colonies with smooth edges and a high nuclear-to-cytoplasmic ratio. Functional pluripotency was confirmed by directed trilineage differentiation followed by immunocytochemical analysis (Figure 13).

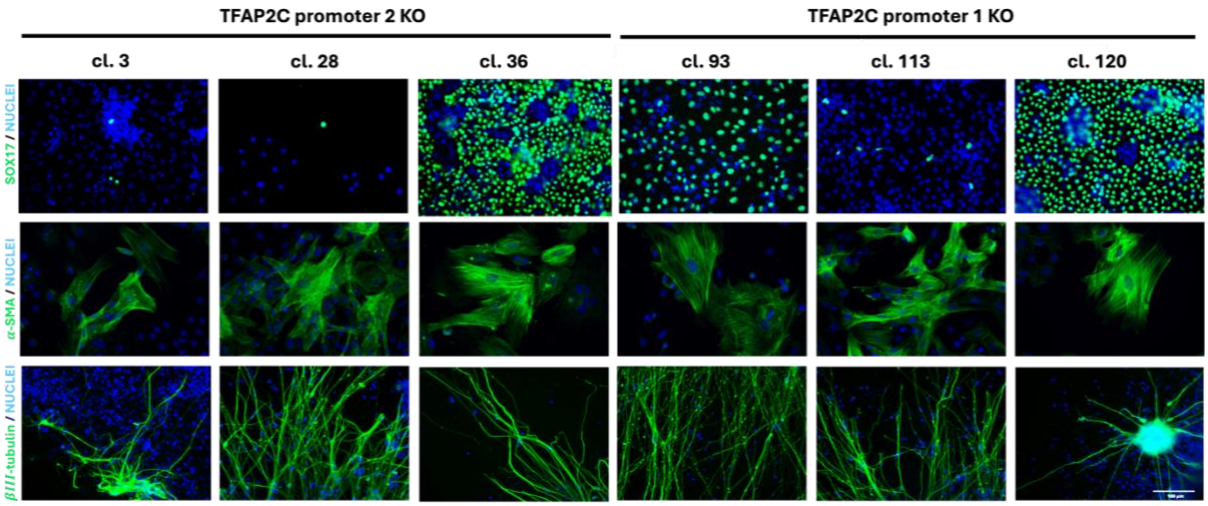


Figure 13. Representative immunocytochemistry images demonstrating successful differentiation into all three germ layers. *Endoderm shows nuclear SOX17 expression at day 5. Mesodermal cultures display organized α-smooth muscle actin (α-SMA) filament networks at day 24. Ectodermal differentiation produces β-III-tubulin-positive neurite structures by day 24, indicating neural phenotype. Nuclei are counterstained with Hoechst (blue). Scale bar: 100 µm.*

All TFAP2C knockout clones were capable of generating derivatives of the three germ layers, as indicated by lineage-specific marker expression and associated morphological changes. Under E6 conditions, neural induction produced dense networks of β-III-tubulin-positive cells by day 24. These cultures exhibited clear filamentous projections typical of the neuronal ectoderm.

Endoderm differentiation induced SOX17 expression by day 5, although intensity and uniformity of SOX17 staining varied among clones. Several clones exhibited strong nuclear SOX17 in the majority of cells, whereas clones 3, 28 and 113 showed reduced or heterogeneous SOX17 expression, indicating variable efficiency of endoderm induction. While TFAP2C promoter region deletion effects cannot be excluded, sensitivity of endoderm differentiation to culture conditions and clonal variation may also contribute.

Mesoderm differentiation produced α-smooth muscle actin-positive cells by day 24. These cells formed dense cellular layers and multicellular assemblies characterized by prominent actin fiber organization. α-smooth muscle actin staining was detected in bundled and network-like cytoskeletal structures, consistent with mesodermal differentiation.

Overall, these results indicate that the analysed TFAP2C knockout clones retained the ability to generate derivatives of all three embryonic germ layers and did not show a general loss of pluripotency-associated developmental potential. Phenotypes observed in downstream experiments can be interpreted in the context of TFAP2C variant loss and are less likely to reflect a gross loss of differentiation competence.

3.2 Each TFAP2C promoter independently drives gene expression when activated by CRISPRa

After verifying that all edited clones maintained pluripotency and differentiation capacity, TFAP2C expression was induced using CRISPR activation. This allowed assessing whether targeting individual promoters with CRISPRa could drive TFAP2C expression on its own and to examine promoter-specific transcriptional regulation in a controlled manner. Single-guide RNAs (sgRNAs) were designed to target either the P1 or the upstream P2 promoter.

The experiments included three genotypes: P1-KO, P2-KO, and P1&P2-wild-type, together with three CRISPRa conditions: a no-guide control, P1-targeting gRNAs, and P2-targeting gRNAs. Three biological replicates were analysed for each genotype. For P1-KO and P2-KO, these replicates were three independent clonal lines, whereas wild-type replicates were obtained from three separate passages. For each knockout genotype, three independent parental clones were analysed. Separate stable P1- and P2-targeting CRISPRa derivative populations were generated from each parental clone by PiggyBac transfection and puromycin selection, and were compared with the corresponding no-guide parental clone. CRISPRa conditions were matched by parental clone but represented independently generated stable cell populations. All CRISPRa experiments were run in parallel, allowing comparison of promoter- and genotype-specific transcriptional responses during early differentiation at days 2, 3, 5 and 7 in E6 medium.

3.2.1 TFAP2C protein expression requires the targeted promoter to be intact

To assess whether activation of either *TFAP2C* promoter region alone is sufficient to induce protein expression, TFAP2C immunostaining was analysed on day 3 as a measure of promoter activity (Figure 14). TFAP2C expression patterns differed among wild-type, P1-KO, and P2-KO cell lines. Across all three genotypes, no-guide controls showed very low TFAP2C signal. TFAP2C-positive cells were rarely detected (Figure 14B). These observations indicate that early differentiation in the absence of CRISPRa showed no appreciable TFAP2C staining and therefore provide a baseline for evaluating CRISPRa-mediated induction.

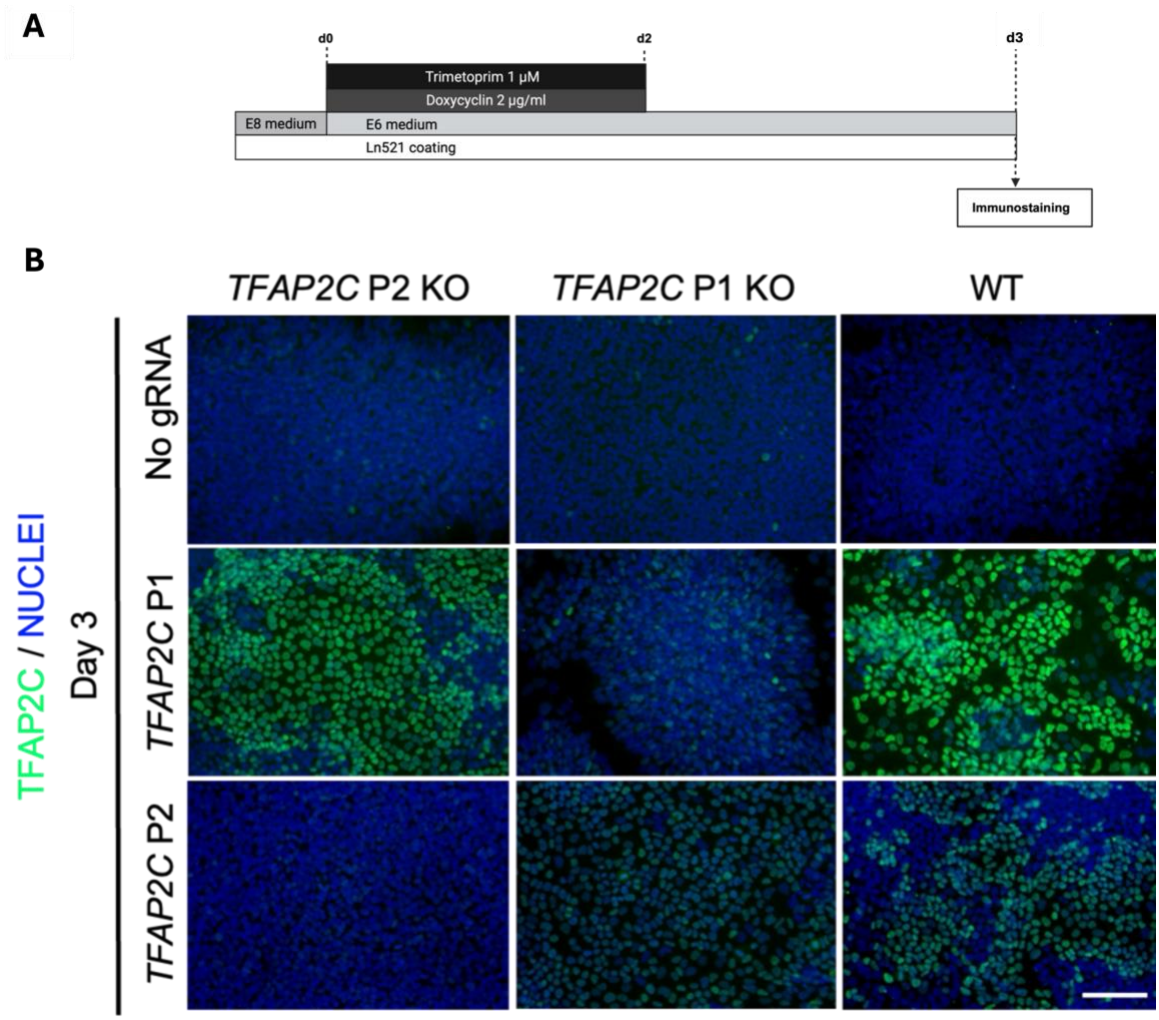


Figure 14. Experimental setup and TFAP2C expression following CRISPR activation. ***(A)*** *Overview of experimental workflow.* ***(B)*** *Representative immunofluorescence images showing TFAP2C expression under the indicated treatment period. Scale bar: 100 µm.*

Following CRISPRa targeting of the P2 promoter, TFAP2C protein was detected in both wild-type and P1-KO cells (Figure 14B), and representative images showed widespread nuclear TFAP2C staining. Immunofluorescence intensity was similar across cells in the well, with only small differences observed. These small differences could reflect variability in CRISPRa efficiency across the population rather than true mosaic expression. Such variability may arise from differences in trimethoprim-dependent stabilization of the DD domain in the dCas9-VP192 construct, as well as variable activity of the doxycycline-inducible rtTA system. The clear induction observed in P1-KO cells demonstrates that activation of the P2 promoter alone is sufficient to drive TFAP2C expression, independent of P1.

CRISPRa targeting of the P1 promoter showed a complementary pattern, resulting in robust TFAP2C protein expression in both WT and P2-KO cells (Figure 14B). This result confirms that either promoter can support TFAP2C protein accumulation when intact and directly activated. Targeting a region upstream of the deleted region did not yield strong TFAP2C induction, although a slight increase in signal was observed compared with no-guide controls.

The lack of a strong response in promoter-deleted genotypes indicates that full TFAP2C activation depends on the presence of the targeted promoter. The weak signal that was detected may result from low-level CRISPRa cross-activation between promoter regions. However, because immunostaining is inherently semi-quantitative, small changes in protein abundance cannot be reliably resolved.

3.2.2 Quantitative analysis confirms promoter specificity and detects weak cross-activation

Promoter-resolved RT-qPCR was used to assess TFAP2C transcription from the alternative P1 and P2 promoters following CRISPR activation targeted to P1 or P2. Forward primers targeted the alternative exon 1 regions of P1- or P2-derived transcripts, while a shared reverse primer in exon 2 detected properly spliced transcripts. RNA samples were collected on days 0, 2, 3, 5, and 7 of E6 differentiation (Figure 15A). Expression values were compared with WT no-gRNA on day 0, and statistical comparisons, marked by asterisks, compared each CRISPRa condition with the matched no-guide control from the same genotype.

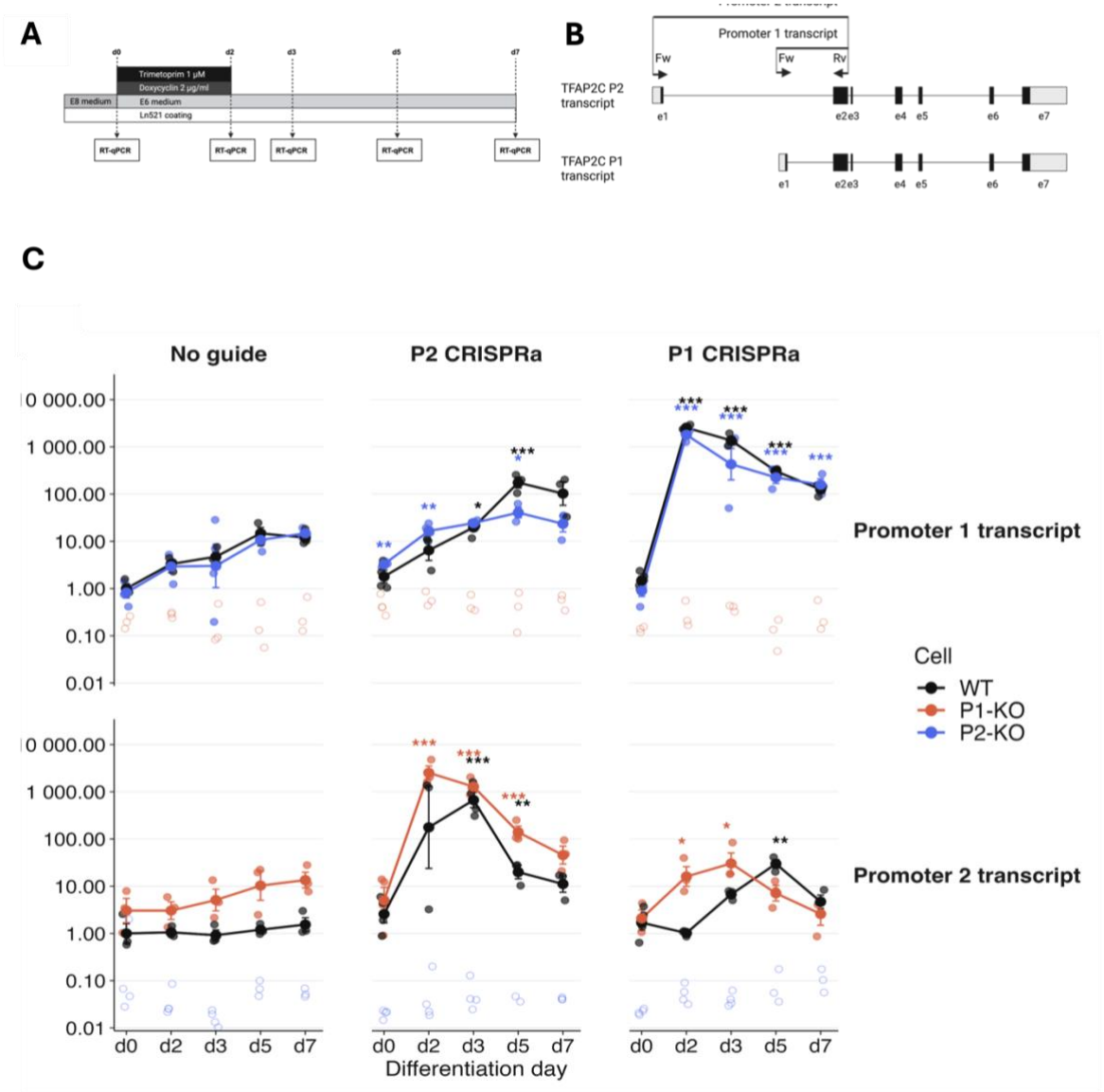


Figure 15. Promoter-resolved quantification of TFAP2C transcription during early differentiation in WT, P1-KO and P2-KO hESCs. ***(A)*** *Schematic of the differentiation and sampling workflow. Cells were switched from E8 to E6 medium on Ln-521 and maintained under doxycycline (2 µg/ml) and trimethoprim (1 µM) to activate the dCas9-VP192 CRISPRa system. RNA was collected at days 0, 2, 3, 5 and 7 for promoter-resolved RT-qPCR analysis.* ***(B)*** *Structure of TFAP2C transcripts and locations of promoter-specific and total TFAP2C primer sets. The P2-specific forward primer binds within exon 1 of the P2 transcript, and the P1-specific forward primer binds within exon 1 of the P1 transcript. Both promoter assays use a shared reverse primer in exon 2 to detect correctly spliced isoforms. Exons are shown as boxes, untranslated regions are white, and coding regions are dark grey.* ***(C-D)*** *Promoter-resolved TFAP2C expression dynamics across differentiation. mRNA abundance was normalised to PPIG and expressed relative to wild-type hESCs at day 0 using the 2^-ΔΔCt method. Data are from n = 3 independent biological replicates per condition, and error bars show SEM. Statistical comparisons were performed using two-sided tests. P-values were adjusted for multiple testing using the Benjamini-Hochberg false-discovery-rate procedure. Significance is indicated according to adjusted q-values: *q < 0.05, **q < 0.01, ***q < 0.001. Wild-type cells (black), P1-KO (red) and P2-KO (blue) were analysed in parallel.* ***(C)*** *P2-derived transcript levels. WT and P1-KO cells exhibit a strong induction peak at day 3, whereas P2-KO cells lack detectable intact P2-derived transcription but retain low residual assay signal, which was treated as background and was not subjected to primary testing.* ***(D)*** *P1-derived transcript levels. WT and P2-KO cells show a sharp induction peak at day 3, whereas P1-KO cells lack detectable intact P1-derived transcription and display only faint CRISPRa-dependent background activity.*

CRISPRa targeting of the P1 promoter strongly induced P1-derived transcript levels when the P1 promoter and its associated exon 1 were intact (Figure 15C-D). In P2-KO cells, which retain the P1 promoter, P1 CRISPRa induced P1-derived transcripts by 836.7-fold at day 2 ($q = 5.42 \times 10^{-13}$), 907.5-fold at day 3 ($q = 5.28 \times 10^{-13}$), 21.1-fold at day 5 ($q = 2.92 \times 10^{-6}$), and 10.7-fold at day 7 ($q = 9.64 \times 10^{-5}$), compared with matched P2-KO no-guide control from the same day. WT cells showed the same pattern, P1 CRISPRa produced 769.8-fold induction relative to the WT no-guide control at day 2 ($q = 3.80 \times 10^{-11}$) and 290.4-fold at day 3 ($q = 8.38 \times 10^{-10}$), and 20.7-fold at day 5 ($q = 4.14 \times 10^{-5}$). This was consistent with the strong increase in TFAP2C protein observed in immunofluorescence staining of promoter-intact cells. P1-KO cells showed no detectable P1-derived transcripts.

P2-directed CRISPRa produced analogous results for the P2-derived transcript. In P1-KO cells, where the P2 promoter remains intact, P2 CRISPRa induced the P2 transcript by 828.8-fold at day 2 ($q = 1.66 \times 10^{-10}$), 252.0-fold at day 3 ($q = 7.87 \times 10^{-9}$), and 13.5-fold at day 5 ($q = 7.19 \times 10^{-4}$). The day-7 effect in P1-KO cells was weaker and did not remain significant after FDR correction (3.38-fold, $q = 0.089$). In WT cells, P2 CRISPRa produced strong early P2-derived transcript levels, peaking at day 3, 943.0-fold, $q = 1.44 \times 10^{-7}$). In P2-KO cells, the intact P2-derived exon 1 – exon 2 transcript was not detected with the promoter-specific assay.

Weak secondary cross-promoter activation was also detectable but weaker than the on-target induction. In WT cells, P1 CRISPRa produced a small P2-transcript expression at day 3 that narrowly missed FDR-adjusted significance (7.3-fold, $q = 0.053$), followed by a significant effect at day 5 (24.8-fold, $q = 0.003$), whereas the day-7 response was not significant (1.5-fold, $q = 0.707$). Conversely, P2-directed CRISPRa increased the P1-derived transcript 4.1-fold at day 3 ($q = 0.034$), 11.9-fold at day 5 ($q = 5.11 \times 10^{-4}$), and 5.1-fold at day 7 ($q = 0.016$). These cross-promoter responses were substantially smaller and delayed than the strongest on-target promoter responses.

Together, these data show that either promoter region can independently drive promoter-specific *TFAP2C* transcription when intact and selectively activated. Deletion of the corresponding promoter-associated region eliminated detection of the intact promoter-specific transcript with these assays, while low-level weaker secondary transcription was still observed from the remaining promoter.

3.3 Promoter-directed TFAP2C activation induces a GATA2/KRT7-associated trophoblast-like phenotype

3.3.1 TFAP2C CRISPRa generates GATA2+/KRT7+ epithelial-like colonies

To determine whether activation of *TFAP2C* via distinct promoter elements is sufficient to initiate trophoblast-like differentiation, CRISPRa was used to selectively target either the P1 or P2 promoter in wild-type P1-KO and P2-KO H9 SBact hESC lines, which were maintained in E6 basal medium. Cellular identity was assessed by immunofluorescence staining for the trophoblast-associated markers KRT7 and GATA2, alongside RT-qPCR quantification of *KRT7* and *GATA2* transcripts (Figure 16).

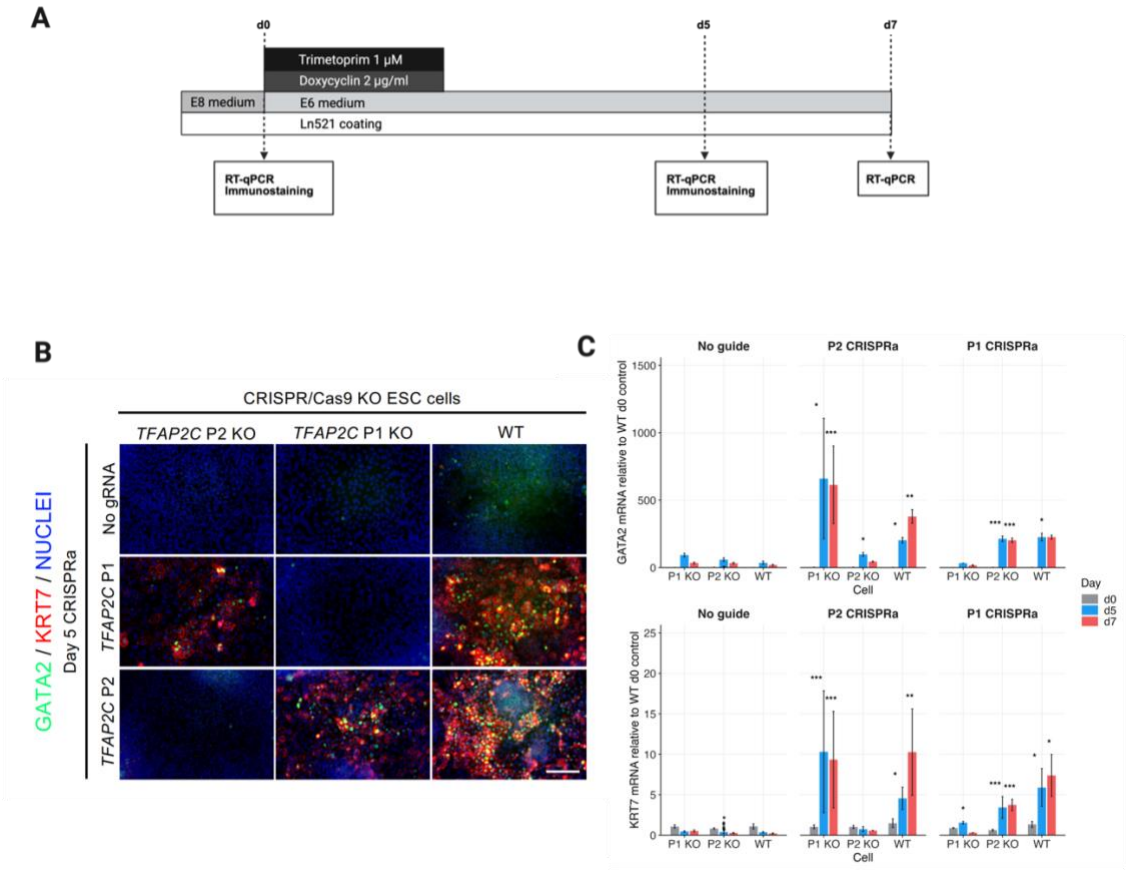


Figure 16. Promoter-specific CRISPR activation induces trophoblast-associated markers in TFAP2C promoter knockout cells. ***(A)*** *Experimental design showing E6 induction with doxycycline (2 µg/mL) and trimethoprim (1 µM) , followed by RT-qPCR and immunostaining at the indicated time points.* ***(B)*** *Representative immunofluorescence staining for GATA2 (green), KRT7 (red), and nuclei (blue) in P1 KO, P2 KO, and WT cells under No guide, P2 CRISPRa, and P1 CRISPRa conditions. Scale: 100 µm.* ***(C)*** *RT-qPCR analysis of GATA2 and KRT7 expression at days 0, 5 and 7. P2 CRISPRa induced marker expression mainly in P1 KO and WT cells, whereas P1 CRISPRa induced expression mainly in P2 KO and WT cells, consistent with promoter-specific TFAP2C activation. Data are from n = 2-3 independent biological replicates per condition, n = 3 except for the failed GATA2 d5 replicate. Error bars show SEM. Statistical comparisons were performed using two-sided tests. P-values were adjusted for multiple testing using the Benjamini-Hochberg false-discovery-rate procedure. Significance is indicated as *q < 0.05, **q < 0.01, ***q < 0.001.*

In the absence of guide RNAs, WT, P1-KO, and P2-KO cells cultured in E6 basal medium showed little evidence of trophoblast-like differentiation, as indicated by GATA2 and KRT7 staining. In contrast, when P2 was activated with CRISPRa, both WT and P1-KO cells formed large, uniform, epithelial-like colonies. These colonies exhibited strong nuclear GATA2 staining and robust cytoplasmic KRT7 staining (Figure 16, Panel B), suggesting coordinated induction of a trophoblast-associated programme. Similarly, when P1 was activated, both WT and P2-KO cells showed strong nuclear GATA2 staining and cytoplasmic KRT7 staining.

This morphological change was accompanied by increased GATA2 and KRT7 transcript abundance in the same cell lines at days 5 and 7, as detected by RT-qPCR (Figure 16C). Statistical testing was performed by comparing each CRISPRa condition with the matched no-guide control from the same genotype and time point. P2-directed CRISPRa induced marker transcripts in P1-KO cells, which retain the P2 promoter. *GATA2* increased by 4.55-fold at day 5 ($q = 0.019$) and 15.49-fold at day 7 ($q = 1.60 \times 10^{-4}$). *KRT7* increased by 12.40-fold at day 5 ($q = 5.89 \times 10^{-4}$) and 12.62-fold at day 7 ($q = 5.89 \times 10^{-4}$). Each comparison was made against the matched P1-KO no-guide control from the same day. WT cells showed similar overall responses, although not every marker/time point remained significant after FDR correction.

Conversely, a similar response was observed after P1-targeted CRISPRa. In P2-KO cells, where the P1 promoter remains intact, P1 CRISPRa increased *GATA2* by 3.8-fold at day 5 ($q = 4.4 \times 10^{-5}$) and 6.3-fold at day 7 ($q = 1.52 \times 10^{-6}$). *KRT7* also increased by 8.6-fold at day 5 ($q = 5.89 \times 10^{-4}$) and 14.5-fold at day 7 ($q = 9.33 \times 10^{-5}$). WT cells showed similar overall directional responses, although not every marker/time-point comparison remained significant after FDR correction.

The corresponding deleted-target background showed weaker responses. In P2-KO cells treated with P2 CRISPRa, GATA2 and KRT7 increased only 1.36-fold and 2.32-fold, respectively, and neither effect reached significance. In P1-KO cells treated with P1 CRISPRa, GATA2 and KRT7 were not induced at day 7 (0.48-fold and 0.61-fold), respectively.

Direct baseline-adjusted promoter-dependence tests further supported P1 dependence. On day 7, the P1-directed response was significantly higher in the P1-intact P2-KO background than in P1-KO cells for both *GATA2* (15.7-fold response ratio,

q = 0.034) and *KRT7* (20-fold, q = 0.034). P2-directed responses were also higher in the P2-intact P1-KO background than in P2-KO cells, but these direct genotype contrasts did not remain significant after FDR correction.

Taken together, strong trophoblast-associated marker induction was most consistently observed when CRISPRa targeted a promoter-associated region retained in the corresponding genotype, with either P1 or P2 sufficient to drive the morphological and transcriptional features of TFAP2C-mediated trophoblast-like fate conversion. Deletion of the corresponding promoter-associated region markedly attenuated the response to the promoter-directed CRISPRa. Concordance between immunofluorescence and RT-qPCR analyses further supports apparent functional redundancy under enforced CRISPRa conditions.

4 Discussion

4.1 TFAP2C promoter 2 is a functional promoter

A central aim of this thesis was to determine whether the upstream transcriptional signal detected in human preimplantation embryo datasets, such as Töhönen et al. (2015), represents a *bona fide* promoter capable of initiating *TFAP2C* transcription, rather than low-coverage single-cell sequencing noise, enhancer-associated bidirectional transcription typical of zygotic genome activation, or transcription confined to a long non-coding RNA region. The CRISPR activation and knockout data support the conclusion that this upstream element, P2, fulfils key functional criteria for a promoter.

First, CRISPRa targeting of P2 in P1-knockout hESCs induced a 252-fold increase in *TFAP2C* mRNA at day 3 relative to the matched no-guide control and produced robust TFAP2C protein expression. A P2-associated transcript containing the alternative exon 1 spliced to exon 2 was detected, as expected, yielding an alternative transcript variant detectable by transcript-specific RT-qPCR. This combination of strong, directional transcription, productive splicing into the canonical coding sequence, and subsequent protein production is more consistent with authentic promoter activity than with non-productive enhancer RNA transcription, which is usually unspliced, and unstable.

Second, deletion of P2 specifically abolished the detectable intact P2-derived *TFAP2C* transcript variant in induced cells, whereas transcription from the canonical P1 promoter remained unchanged. Because the P2-knockout was N-terminally targeted, these data do not exclude the possibility that partial transcription was still initiated from P2 and persisted partially upstream or downstream of the deleted segment. The result therefore shows that P2 is required for expression of the full upstream transcript variant detected in this system, while still supporting the conclusion that the observed signal is not a technical artefact or transcriptional noise.

The TFAP2C locus also produces a P2-associated long non-coding RNA. Many developmental genes utilise alternative promoters to generate lncRNAs that regulate chromatin accessibility, enhancer-promoter communication, or three-dimensional genome organisation at the same locus (Engreitz et al. 2016; Kopp et al. 2018, Qin et al. 2015). Given that TFAP2C has pioneer-like activity, enabling access to regulatory

DNA and priming cis-regulatory elements in naïve hESCs and trophoblast stem cells, it is plausible that the P2-associated lncRNA, or the activity of the P2 locus itself, could help establish or maintain a permissive chromatin environment during cleavage-stage development. More broadly, this suggests that promoter-derived transcription at the TFAP2C locus may participate in stage-specific regulation of promoter usage, including the balance between P1 and P2 usage.

4.1.1 Promoter compensation under enforced TFAP2C activation may mask context-dependent promoter function

Deleting either the P1 or P2 promoter region had only a limited effect on CRISPRa-induced trophoblast-like differentiation when the remaining promoter was activated. In conditions where the remaining promoter was activated, TFAP2C expression was maintained, and colonies expressing KRT7 and GATA2 were consistently observed. In this experimental context, the ability to initiate a trophoblast-like transcriptional programme therefore appears to depend primarily on the total TFAP2C output rather than on which promoter was used.

This interpretation aligns with evidence that TFAP2C functions in a dosage-sensitive manner across multiple stages of trophoblast differentiation. In the preimplantation embryo, TFAP2C acts upstream of *Cdx2* and contributes to cell polarity and position-dependent Hippo signalling during trophoblast specification. TFAP2C depletion disrupts apical polarity, activates Hippo signalling in outer blastomeres, and reduces expression of regulators linked to actin organisation (Cao et al. 2015; Kuckenberg et al. 2012; Cosemans et al. 2023). More recent work further places TFAP2C within a TEAD-linked bistable network that promotes lineage diversification and regulates Hippo pathway components before lineage segregation (Zhu et al. 2024).

Evidence from later trophoblast contexts likewise supports sensitivity to TFAP2C levels and timing. TFAP2C is necessary to maintain trophoblast stem cells (Varberg et al. 2023; Kim et al. 2024; Kuckenberg et al. 2012). In mice, reduced Tfap2c dosage impairs placental trophoblast differentiation, lowers stem-associated markers including Cdx2, Eomes, Gata3, and Ascl2, and alters maternal blood spaces (Kaiser et al. 2015). In human trophoblast stem cells, TFAP2C is critical for stem-state maintenance and the onset of EVT differentiation, but must subsequently decline for appropriate late EVT gene activation, indicating that both presence and level are

developmentally staged variables rather than static requirements (Varberg et al. 2023; Kim et al. 2024).

TFAP2C function is shaped by Hippo signalling, interactions with TEAD4 and KLF5, maternal retinoic acid metabolism, promoter-enhancer interactions, and chromatin constraints, including H3K9me3 (Cao et al. 2015; Zhu et al. 2024; Gao et al. 2024; Wang et al. 2024). TFAP2C also binds promoters of early ICM- and TE-associated genes during preimplantation development, reinforcing the idea that its output and genomic action are context-dependent rather than simply lineage-restricted.

Against that background, strong CRISPRa with VP192 is likely to bypass some endogenous constraints by driving transcription above the threshold required to initiate the trophoblast programme. Under those conditions, activation through either remaining promoter would be expected to preserve the phenotype, even if P1 and P2 normally differ in timing or signalling responsiveness. The present findings therefore support apparent functional compensation under enforced activation, but do not demonstrate physiological equivalence of the two promoters *in vivo*.

4.2 TFAP2C isoforms show quantitative differences in hESCs, but may diverge functionally *in vivo*

Promoter-specific knockouts did not reveal consistent functional differences between TFAP2C isoforms in the hESC CRISPRa system. Although some experiments suggested that P2 targeting produced more consistent KRT7- and GATA2-positive trophoblast-like colonies, whereas P1-targeted cells showed stronger TFAP2C immunofluorescence, these trends were not reproducible across replicates or culture conditions. In this system, total TFAP2C dosage appears to matter more than isoform identity, consistent with previous evidence that TFAP2C overexpression can broadly increase enhancer accessibility and support cell-state transitions (Pastor et al. 2018; Li et al. 2024; Kim et al. 2024; Theurillat et al. 2020).

A more cautious *in vivo* model is nevertheless plausible. Because P1 and P2 encode distinct N termini, promoter choice may alter isoform function, not just transcript abundance. Notably, only the P1 isoform contains a SUMO acceptor lysine at position 10, raising the possibility that it engages more readily with SUMO-associated repressive chromatin environments.

TFAP2C cofactors and targets also change across developmental stages. In early embryos, TFAP2C acts with NR5A2 at shared enhancers linked to inner cell mass and trophectoderm programmes. By blastocyst and peri-implantation stages, however, its binding profile shifts toward late trophectoderm and extraembryonic genes, with greater cooperation with factors such as CDX2 and SOX2. In human blastocyst models and BAP-treated pluripotent stem cells, TFAP2C also works with GATA2/3 and TFAP2A to promote trophectoderm identity while repressing pluripotency genes.

Within this framework, the SUMOylatable P1 isoform may be better suited to extraembryonic ectoderm or trophoblast stem cell-like contexts, where chromatin is more compact and repressive complexes are present, including high H3K9me3 and abundant SUMO-modified PRC1.6/SETDN1/KAP1/HDAC complexes (Theurillat et al. 2020; Wang et al. 2024; Cossec et al. 2018). Conversely, the non-SUMOylated P2 isoform may be more compatible with totipotent blastomeres or naïve epiblast-like states, where chromatin is more accessible and TFAP2C functions primarily as an activating or pioneer factor.

CRISPR activation in hESCs likely masks these qualitative differences. SUMO conjugation is constrained by the limited availability of SUMOylation-linked enzymes such as UBC9 and E3 activity. System saturation can cause chromatin decompaction and loss of H3K9me3 (Theurillat et al. 2020; Cossec et al. 2018). Strong TFAP2C overexpression can override endogenous chromatin constraints and activate trophoblast-associated programmes before normal lineage-restricted networks are established (Kim et al. 2024; Pastor et al. 2018; Krendl et al. 2017). Under these non-physiological conditions, both isoforms may bind broadly, independent of SUMO status, and produce similar phenotypes despite potentially distinct functions in embryos. Overall, the hESC CRISPRa data support a mainly dosage-dependent effect i*n vitro*, while still leaving open the possibility of context-dependent isoform specialisation *in vivo* (Wang et al. 2024; Li et al. 2023; Krendl et al. 2017; Driscoll et al. 2024).

4.3 Promoter usage provides a mechanism for developmental diversification of TFAP2C function

TFAP2C is a highly adaptive lineage regulator whose activity is shaped by chromatin state, partner transcription factors, and cis-regulatory architecture. During development, it supports multiple, seemingly opposing cell states (Li et al. 2024; Wang et al. 2024; Pastor et al. 2018). At the 2-8-cell totipotent stages, TFAP2C acts as a bipotency factor by co-activating inner cell mass and trophectoderm gene programmes (Li et al. 2024; Zhu et al. 2024).

TFAP2C also primes naïve pluripotency and OCT4/SOX2 regulatory elements in naïve hESCs and germline-like cells, while remaining essential for establishing and maintaining trophoblast stem cells in mouse and human (Kuckenberg et al. 2012; Pastor et al. 2018). At later stages, TFAP2C contributes to extravillous trophoblast invasion, deep placentation, and surface ectoderm/epidermal specification often through pioneer-like enhancer priming (Li et al. 2019; Kim et al. 2024; Varberg et al. 2023). Similar stage-specific redeployment occurs in other TFAP2 paralogs, such as the TFAP2A-to-TFAP2B switch during neural crest development (Rothstein and Simoes-Costa, 2020).

Developmentally regulated promoter usage may provide a mechanism for encoding TFAP2C context specificity from a single locus. Single-cell 5'-end transcriptomic studies in human oocytes and cleavage-stage embryos suggest widespread alternative promoter usage during zygotic genome activation, including at TFAP2C and multiple PRD-like homeobox transcription factors. During this early window, TFAP2C transcription appears biased toward the P2 promoter, consistent with its bipotent role and its binding to broadly accessible promoters of both inner cell mass and trophectoderm genes (Töhönen et al. 2015).

By the blastocyst and peri-implantation stages, the regulatory landscape becomes more restricted, marked by increased H3K9me3 deposition, altered enhancer structure, and Hippo-dependent polarity cues. In this more restricted landscape, early embryo data indicate that TFAP2C transcription shifts towards the P1 promoter (Töhönen et al. 2015), producing a SUMO-competent isoform that may be better suited to lineage-restricted chromatin environments and stage-specific cofactor networks.

More broadly, alternative promoter usage is a well-established mechanism of developmental regulation. Genome-wide 5'-end profiling approaches (GAGE/NET-CAGE/RAMPAGE), together with organ-development datasets, show that a large fraction of genes use stage- and tissue-specific promoters. These promoters independently regulate N-terminal structure, mRNA fate, or expression dosage (Batut et al. 2013; Tan et al. 2025; Ayoubi et al. 1996). Many genes traditionally classified as "housekeeping" similarly acquire cell-type-specific functions through alternative transcription start sites during development (Tan et al. 2025; Ayoubi et al. 1996; Batut et al. 2013; Li et al. 2019).

These observations support a model in which promoter selection and isoform regulation are active, regulated processes that shape the pleiotropic functions of TFAP2C. The early P2-biased isoform and the later P1-biased SUMO-competent isoform align with a broader strategy in which alternative promoters, chromatin state, 3D architecture, and cofactor networks (NANOG, CDX2, TEAD4, DLX6, EPAS1) together determine how a single factor drives divergent lineage trajectories during early embryo development (Wang et al. 2024).

A deeper understanding of this regulation may also improve *in vitro* models of extraembryonic development. Naïve human pluripotent stem cells are generally considered more competent than primed pluripotent stem cells at generating trophoblast stem cell-like populations that resemble their *in vivo* counterparts. Trophoblast-like populations derived from primed hESCs may display incomplete or aberrant trophoblast identities, including abnormal expression of lineage-associated markers such as the chromosome 19 microRNA cluster, C19MC.

4.4 Limitations and future directions

Interpreting these findings requires recognising that the current study simplifies the biology by activating *TFAP2C* promoters via CRISPRa in pluripotent stem cells, rather than within the temporal, spatial and signalling context of the embryo. As a result, the roles of promoters in processes like implantation, early differentiation, and extraembryonic tissue formation might not be fully captured. Although P1 and P2 act redundantly under CRISPRa conditions, this does not imply equivalent roles *in vivo*, where promoter use follows specific timing and involves different chromatin states or

regulatory factors at various developmental stages. Additionally, mosaic editing or heterogeneous CRISPRa efficiency can impact data interpretation.

A central limitation is the lack of promoter- or isoform-resolved epigenomic data. Current TFAP2C maps in hESCs and embryos define global enhancer occupancy but do not distinguish between P1- and P2-derived transcripts or protein isoforms. This also limits the ability to determine whether P1- and P2-derived transcripts interact with different enhancer landscapes, recruit different cofactors, or differ in post-translational regulation. Addressing these questions will require promoter-specific chromatin profiling, such as CUT&RUN or CUT&Tag targeting initiation-associated complexes, combined with proteomic analyses comparing SUMOylated and non-SUMOylated TFAP2C.

An additional limitation is the absence of a P1/P2 double-knockout line. Although single promoter region/exon 1 deletions showed a clear reduction in expression, a double deletion would have allowed testing whether *TFAP2C* transcription can originate from an alternative promoter outside the P1 or P2 region. Residual initiation from remaining promoter sequences also cannot be excluded. Genomic integrity and off-target effects were not assessed, and the magnitude of CRISPRa can also be influenced by guide efficiency or integration. The protein isoforms themselves were not directly distinguished, and the contribution of the P2-associated lncRNA remains unresolved because the P2-associated deletion overlaps regulatory sequences associated with both the *TFAP2C* P2 and the annotated lncRNA, making it impossible to separate their individual functions. KO replicates were independent clones, whereas wild-type replicates were separate passages. Therefore, unequal replicate structure and clone-specific effects may contribute to genotype differences.

The study also relies on bulk RT-qPCR, which averages across heterogeneous cell states and obscures potential cell-to-cell differences in promoter usage. Single-cell long-read transcriptomics would be needed to determine whether the P1 and P2 promoters drive exclusive transcriptional programmes or coexist within the same cells across different differentiation states.

The differentiation analysis was based on a very limited set of markers, namely TFAP2C, KRT7, and GATA2. These markers indicate the initiation of an extraembryonic or epithelial transcriptomic programme but are insufficient to

establish a specific trophoblast identity or to distinguish trophoblast-like cells from amniotic or epidermal cell states. Broader transcriptomic profiling is needed to determine the full effect. TFAP2C overexpression has also been used to promote epidermal differentiation from primed hESCs (Li et al. 2019).

Finally, 2D culture systems limit assessment of promoter-specific functions that may depend on tissue architecture and multicellular interactions. Processes such as trophectoderm polarisation, EVT invasion, syncytialisation, and trophoblast maturation are likely influenced by extracellular matrix composition, mechanical cues, maternal cell interactions, and signalling pathways such as Hippo signalling. Recent advances in trophoblast organoids, blastoids, and endometrial co-culture and assembloid models would enable the study of these processes in more physiologically relevant settings (Molè et al. 2026; Turco et al. 2018).

Overall, resolving *TFAP2C* promoter function will require integrated single-cell and isoform-resolved approaches, combined with promoter-specific chromatin mapping and proteomic analysis in more physiologically relevant developmental models.

4.5 Conclusion

This thesis provides functional evidence that P2 is a *bona fide* promoter capable of initiating *TFAP2C* transcription independently of P1. CRISPRa targeting either P1 or P2 induced promoter-associated *TFAP2C* transcripts and protein in genetic backgrounds that retained the corresponding regulatory region. Under enforced CRISPRa conditions, activation of either promoter was sufficient to induce a GATA2/KRT7-associated epithelial trophoblast-like phenotype, with no reproducible evidence for distinct promoter-specific cell states in the measured GATA2/KRT7-associated phenotype. These findings support apparent functional redundancy under enforced TFAP2C activation, while leaving open the possibility that P1 and P2 have temporally or contextually distinct functions during human development. Future promoter-resolved transcriptomic, chromatin and protein-level analyses will be required to resolve these physiological differences.

5 Data and code availability

Raw data, processed data, analysis code, and any additional information required to reanalyse the data reported in this thesis are available upon request from the author, Viljami Wuollet (viljami.wuollet@helsinki.fi).

6 Acknowledgements

First, I would like to thank my supervisor, Jere Weltner, for all the information, for the education in the wonderful world of stem cells, and for the patience shown throughout this project. Thank you, Juha Kere, for warmly welcoming me to your research group and for the informal, friendly discussions, for example on the sauna benches.

I would also like to thank Auli Saarinen for her help with pipetting in the lab, Sini Ezer for her help with the equipment, and the whole Juha Kere lab in Helsinki and Stockholm, especially Tuuli Vallo for her peer support on this massive TFAP2C project. I'm keeping all my fingers crossed for the Vallo et al. 2026 paper!

And above all, my deepest thanks go to my beloved wife Raisa and my family for their support and love.

This work was supported by the Sakari Alhopuro Foundation, the Sigrid Jusélius Foundation, the Jane and Aatos Erkko Foundation, and Finska Läkaresällskapet.

7 Author contribution

Study design (Viljami Wuollet, Jere Weltner), laboratory work, sample preparation, programming, statistical analysis, visualisation, literature review, interpretation, and writing (Viljami Wuollet).

8 References

Ayoubi, T. A. Y. and Van de Ven, W. J. M. (1996) 'Regulation of gene expression by alternative promoters', The FASEB Journal, 10(4), pp. 453-460. https://doi.org/10.1096/fasebj.10.4.8647344.

Balboa, D., Weltner, J., Eurola, S., Trokovic, R., Wartiovaara, K. and Otonkoski, T. (2015) 'Conditionally Stabilized dCas9 Activator for Controlling Gene Expression in Human Cell Reprogramming and Differentiation', Stem Cell Reports, 5(3), pp. 448-459. https://doi.org/10.1016/j.stemcr.2015.08.001.

Balestrini, P. A., Abdelbaki, A., McCarthy, A., Devito, L., Senner, C. E., Chen, A. E. et al. (2024) 'Transcription factor-based transdifferentiation of human embryonic to trophoblast stem cells', Development, 151(17), article dev202778. https://doi.org/10.1242/dev.202778.

Batut, P., Dobin, A., Plessy, C., Carninci, P. and Gingeras, T. R. (2013) 'High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression', Genome Research, 23(1), pp. 169-180. https://doi.org/10.1101/gr.139618.112.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', Journal of the Royal Statistical Society: Series B (Methodological), 57(1), pp. 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Cao, Z., Carey, T. S., Ganguly, A., Wilson, C. A., Paul, S. and Knott, J. G. (2015) 'Transcription factor AP-2γ induces early Cdx2 expression and represses HIPPO signaling to specify the trophectoderm lineage', Development, 142, pp. 1606-1615. https://doi.org/10.1242/dev.120238.

Cartwright, J. E., Kenny, L. C., Dash, P. R., Crocker, I. P., Aplin, J. D., Baker, P. N. and Whitley, G. S. J. (2002) 'Trophoblast Invasion of Spiral Arteries: a Novel In Vitro Model', Placenta, 23(2-3), pp. 232-235. https://doi.org/10.1053/plac.2001.0760.

Castel, G., Meistermann, D., Bretin, B., Firmin, J., Blin, J., Loubersac, S. et al. (2020) 'Induction of Human Trophoblast Stem Cells from Somatic Cells and Pluripotent Stem Cells', Cell Reports, 33(8), article 108419. https://doi.org/10.1016/j.celrep.2020.108419.

Chen, Y., Siriwardena, D., Penfold, C., Pavlinek, A. and Boroviak, T. E. (2022) 'An integrated atlas of human placental development delineates essential regulators of trophoblast stem cells', Development, 149(13), article dev200171. https://doi.org/10.1242/dev.200171.

Cosemans, G., Boel, A., Bekaert, B., Stoop, D., Chuva de Sousa Lopes, S. M., Coucke, P., Menten, B., Lluís Viñas, F. and Heindryckx, B. (2023) 'P-802 TFAP2C is required for proper segregation of developmental markers regulating trophectoderm commitment during mouse preimplantation development', Human Reproduction, 38(Supplement 1), article dead093.1108. https://doi.org/10.1093/humrep/dead093.1108.

Cossec, J. C., Theurillat, I., Chica, C., Búa Aguín, S., Gaume, X., Andrieux, A. et al. (2018) 'SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States', Cell Stem Cell, 23(5), pp. 742-757.e8. https://doi.org/10.1016/j.stem.2018.10.001.

Derisoud, E., Jiang, H., Zhao, A., Chavatte-Palmer, P. and Deng, Q. (2024) 'Revealing the molecular landscape of human placenta: a systematic review and meta-analysis of single-cell RNA sequencing studies', Human Reproduction Update, 30(4), pp. 410-441. https://doi.org/10.1093/humupd/dmae006.

Ding, X., Yu, L., Chen, L., Li, Y., Zhang, J., Sheng, H. et al. (2022) 'Recent Progress and Future Prospect of CRISPR/Cas-Derived Transcription Activation (CRISPRa) System in Plants', Cells, 11(19), p. 3045. https://doi.org/10.3390/cells11193045.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F. et al. (2016) 'Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9', Nature Biotechnology, 34(2), pp. 184-191. https://doi.org/10.1038/nbt.3437.

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B. L., Xavier, R. J. and Root, D. E. (2014) 'Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation', Nature Biotechnology, 32(12), pp. 1262-1267. https://doi.org/10.1038/nbt.3026.

Dong, C., Beltcheva, M., Gontarz, P., Zhang, B., Popli, P., Fischer, L. A. et al. (2020) 'Derivation of trophoblast stem cells from naïve human pluripotent stem cells', eLife, 9, article e52504. https://doi.org/10.7554/elife.52504.

Driscoll, C. S., Kim, J., Ashry, M. and Knott, J. G. (2024) 'Does TFAP2C govern conflicting cell fates in mouse preimplantation embryos?', Reproduction, 167(4), article e230440. https://doi.org/10.1530/rep-23-0440.

Eloranta, J. J. and Hurst, H. C. (2002) 'Transcription Factor AP-2 Interacts with the SUMO-conjugating Enzyme UBC9 and Is Sumolated in Vivo', Journal of Biological Chemistry, 277(34), pp. 30798-30804. https://doi.org/10.1074/jbc.m202780200.

Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., McDonel, P. E., Guttman, M. and Lander, E. S. (2016) 'Local regulation of gene expression by lncRNA promoters, transcription and splicing', Nature, 539(7629), pp. 452-455. https://doi.org/10.1038/nature20149.

Gao, R., Yang, G., Wang, M., Xiao, J., Yi, S., Huang, Y. et al. (2024) 'Defining a TFAP2C-centered transcription factor network during murine peri-implantation', Developmental Cell, 59(9), pp. 1146-1158.e6. https://doi.org/10.1016/j.devcel.2024.03.015.

Gawriyski, L., Jouhilahti, E. M., Yoshihara, M., Fei, L., Weltner, J., Airenne, T. T. et al. (2023) 'Comprehensive characterization of the embryonic factor LEUTX', iScience, 26(3), article 106172. https://doi.org/10.1016/j.isci.2023.106172.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E. et al. (2013) 'CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes', Cell, 154(2), pp. 442-451. https://doi.org/10.1016/j.cell.2013.06.044.

Gong, X., He, W., Jin, W., Ma, H., Wang, G., Li, J. et al. (2024) 'Disruption of maternal vascular remodeling by a fetal endoretrovirus-derived gene in preeclampsia', Genome Biology, 25(1), article 117. https://doi.org/10.1186/s13059-024-03265-z.

Handwerger, S. (2010) 'New insights into the regulation of human cytotrophoblast cell differentiation', Molecular and Cellular Endocrinology, 323(1), pp. 94-104. https://doi.org/10.1016/j.mce.2009.12.015.

Hsu, P. D., Lander, E. S. and Zhang, F. (2014) 'Development and Applications of CRISPR-Cas9 for Genome Engineering', Cell, 157(6), pp. 1262-1278. https://doi.org/10.1016/j.cell.2014.05.010.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V. et al. (2013) 'DNA targeting specificity of RNA-guided Cas9 nucleases', Nature Biotechnology, 31(9), pp. 827-832. https://doi.org/10.1038/nbt.2647.

Io, S., Kabata, M., Iemura, Y., Semi, K., Morone, N., Minagawa, A. et al. (2021) 'Capturing human trophoblast development with naive pluripotent stem cells in vitro', Cell Stem Cell, 28(6), pp. 1023-1039.e13. https://doi.org/10.1016/j.stem.2021.03.013.

Jang, Y. J., Kim, M., Lee, B. K. and Kim, J. (2022) 'Induction of human trophoblast stem-like cells from primed pluripotent stem cells', Proceedings of the National Academy of Sciences, 119(20), article e2115709119. https://doi.org/10.1073/pnas.2115709119.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012) 'A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity', Science, 337(6096), pp. 816-821. https://doi.org/10.1126/science.1225829.

Jouhilahti, E. M., Madissoon, E., Vesterlund, L., Töhönen, V., Krjutškov, K., Plaza Reyes, A. et al. (2016) 'The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation', Development, 143(19), pp. 3459-3469. https://doi.org/10.1242/dev.134510.

Kaiser, S., Koch, Y., Kühnel, E., Sharma, N., Gellhaus, A., Kuckenberg, P., Schorle, H. and Winterhager, E. (2015) 'Reduced Gene Dosage of Tfap2c Impairs Trophoblast Lineage Differentiation and Alters Maternal Blood Spaces in the Mouse Placenta', Biology of Reproduction, 93(2), p. 31. https://doi.org/10.1095/biolreprod.114.126474.

Kim, J., Driscoll, C. S., Li, L., Wilson, C. A., Xie, W. and Knott, J. G. (2025) 'Deciphering the role of cis-regulatory elements and TFAP2C in the activation of zygotic Sox2 expression in mouse preimplantation embryos', Development, 152(14), article dev204626. https://doi.org/10.1242/dev.204626.

Kim, M., Adu-Gyamfi, E. A., Kim, J. and Lee, B. K. (2023) 'Super-enhancer-associated transcription factors collaboratively regulate trophoblast-active gene expression programs in human trophoblast stem cells', Nucleic Acids Research, 51(8), pp. 3806-3819. https://doi.org/10.1093/nar/gkad215.

Kim, M., Jang, Y. J., Lee, M., Guo, Q., Son, A. J., Kakkad, N. A., Roland, A. B., Lee, B. K. and Kim, J. (2024) 'The transcriptional regulatory network modulating human trophoblast stem cells to extravillous trophoblast differentiation', Nature Communications, 15(1), article 1285. https://doi.org/10.1038/s41467-024-45669-2.

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C. et al. (2015) 'Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex', Nature, 517(7536), pp. 583-588. https://doi.org/10.1038/nature14136.

Kopp, F. and Mendell, J. T. (2018) 'Functional Classification and Experimental Dissection of Long Noncoding RNAs', Cell, 172(3), pp. 393-407. https://doi.org/10.1016/j.cell.2018.01.011.

Krendl, C., Shaposhnikov, D., Rishko, V., Ori, C., Ziegenhain, C., Sass, S. et al. (2017) 'GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency', Proceedings of the National Academy of Sciences, 114(45), pp. E9579-E9588. https://doi.org/10.1073/pnas.1708341114.

Krzywinski, M. and Altman, N. (2014a) 'Analysis of variance and blocking', Nature Methods, 11(7), pp. 699-700. https://doi.org/10.1038/nmeth.3005.

Krzywinski, M. and Altman, N. (2014b) 'Two-factor designs', Nature Methods, 11(12), pp. 1187-1188. https://doi.org/10.1038/nmeth.3180.

Kuckenberg, P., Kubaczka, C. and Schorle, H. (2012) 'The role of transcription factor Tcfap2c/TFAP2C in trophectoderm development', Reproductive BioMedicine Online, 25(1), pp. 12-20. https://doi.org/10.1016/j.rbmo.2012.02.015.

Kumar, B., Navarro, C., Winblad, N., Schell, J. P., Zhao, C., Weltner, J. et al. (2022) 'Polycomb repressive complex 2 shields naïve human pluripotent cells from

trophectoderm differentiation', Nature Cell Biology, 24(6), pp. 845-857. https://doi.org/10.1038/s41556-022-00916-w.

Li, L., Lai, F., Hu, X., Liu, B., Lu, X., Lin, Z. et al. (2023) 'Multifaceted SOX2-chromatin interaction underpins pluripotency progression in early embryos', Science, 382(6676), article eadi5516. https://doi.org/10.1126/science.adi5516.

Li, L., Lai, F., Liu, L., Lu, X., Hu, X., Liu, B. et al. (2024) 'Lineage regulators TFAP2C and NR5A2 function as bipotency activators in totipotent embryos', Nature Structural & Molecular Biology, 31(6), pp. 950-963. https://doi.org/10.1038/s41594-023-01199-x.

Li, L., Wang, Y., Torkelson, J. L., Shankar, G., Pattison, J. M., Zhen, H. H. et al. (2019) 'TFAP2C- and p63-Dependent Networks Sequentially Rearrange Chromatin Landscapes to Drive Human Epidermal Lineage Commitment', Cell Stem Cell, 24(2), pp. 271-284.e8. https://doi.org/10.1016/j.stem.2018.12.012.

Li, Y., Moretto-Zita, M., Soncin, F., Wakeland, A., Wolfe, L., Leon-Garcia, S. et al. (2013) 'BMP4-directed trophoblast differentiation of human embryonic stem cells is mediated through a ΔNp63+ cytotrophoblast stem cell state', Development, 140(19), pp. 3965-3976. https://doi.org/10.1242/dev.092155.

Liu, K., Xiao, Y., Gan, L., Li, W., Zhang, J. and Min, J. (2023) 'Structural basis for specific DNA sequence motif recognition by the TFAP2 transcription factors', Nucleic Acids Research, 51(15), pp. 8270-8282. https://doi.org/10.1093/nar/gkad583.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method', Methods, 25(4), pp. 402-408. https://doi.org/10.1006/meth.2001.1262.

Long, J. S. and Ervin, L. H. (2000) 'Using heteroscedasticity consistent standard errors in the linear regression model', The American Statistician, 54(3), pp. 217-224. https://doi.org/10.1080/00031305.2000.10474549.

Mak, K. H. M., Lam, Y. M. and Ng, R. K. (2021) 'Histone demethylase JMJD2B/KDM4B regulates transcriptional program via distinctive epigenetic targets and protein interactors for the maintenance of trophoblast stem cells', Scientific Reports, 11(1), article 884. https://doi.org/10.1038/s41598-020-79601-7.

Messner, S., Schuermann, D., Altmeyer, M., Kassner, I., Schmidt, D., Schär, P., Müller, S. and Hottiger, M. O. (2009) 'Sumoylation of poly(ADP-ribose) polymerase 1 inhibits its acetylation and restrains transcriptional coactivator function', The FASEB Journal, 23(11), pp. 3978-3989. https://doi.org/10.1096/fj.09-137695.

Molè, M. A., Elderkin, S., Zorzan, I., Penfold, C., Horsley, N., Pokhilko, A. et al. (2026) 'Modeling human embryo implantation in vitro', Cell, 189(1), pp. 87-105.e28. https://doi.org/10.1016/j.cell.2025.10.027.

Ohgushi, M., Honda, K., Takagi, R. and Eiraku, M. (2025) 'Stepwise strategy for generating human trophoblast stem-like cells from embryonic stem cells reveals specific role of TFAP2C in acquiring self-renewability', Scientific Reports, 15(1), article 19040. https://doi.org/10.1038/s41598-025-03830-x.

Okae, H., Toh, H., Sato, T., Hiura, H., Takahashi, S., Shirane, K., Kabayama, Y., Suyama, M., Sasaki, H. and Arima, T. (2018) 'Derivation of Human Trophoblast Stem Cells', Cell Stem Cell, 22(1), pp. 50-63.e6. https://doi.org/10.1016/j.stem.2017.11.004.

Omachi, K. and Miner, J. H. (2022) 'Comparative analysis of dCas9-VP64 variants and multiplexed guide RNAs mediating CRISPR activation', PLOS ONE, 17(6), article e0270008. https://doi.org/10.1371/journal.pone.0270008.

Pastor, W. A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T. J. et al. (2018) 'TFAP2C regulates transcription in human naive pluripotency by opening enhancers', Nature Cell Biology, 20(5), pp. 553-564. https://doi.org/10.1038/s41556-018-0089-0.

Qin, D. and Xu, C. (2015) 'Study strategies for long non-coding RNAs and their roles in regulating gene expression', Cellular and Molecular Biology Letters, 20(2), pp. 323-349. https://doi.org/10.1515/cmble-2015-0021.

Roberts, J. M. and Hubel, C. A. (2009) 'The Two Stage Model of Preeclampsia: Variations on the Theme', Placenta, 30 Suppl A, pp. S32-37. https://doi.org/10.1016/j.placenta.2008.11.009.

Roberts, R. M., Ezashi, T., Sheridan, M. A. and Yang, Y. (2018) 'Specification of trophoblast from embryonic stem cells exposed to BMP4', Biology of Reproduction, 99(1), pp. 212-224. https://doi.org/10.1093/biolre/ioy070.

Rothstein, M. and Simoes-Costa, M. (2020) 'Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification', Genome Research, 30(1), pp. 35-48. https://doi.org/10.1101/gr.249680.119.

Sadu Murari, L. S., Kunkel, S., Shetty, A. V., Bents, A., Bhandary, A. and Rivera-Mulia, J. C. (2025) 'p63: A Master Regulator at the Crossroads Between Development, Senescence, Aging, and Cancer', Cells, 14(1), p. 43. https://doi.org/10.3390/cells14010043.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T. et al. (2012) 'Fiji: an open-source platform for biological-image analysis', Nature Methods, 9(7), pp. 676-682. https://doi.org/10.1038/nmeth.2019.

Schmidt, U., Weigert, M., Broaddus, C. and Myers, G. (2018) 'Cell Detection with Star-Convex Polygons', in Medical Image Computing and Computer Assisted Intervention - MICCAI 2018. Lecture Notes in Computer Science, vol. 11071. Cham: Springer, pp. 265-273. https://doi.org/10.1007/978-3-030-00934-2_30.

Seetharam, A. S., Vu, H. T. H., Choi, S., Khan, T., Sheridan, M. A., Ezashi, T., Roberts, R. M. and Tuteja, G. (2022) 'The product of BMP-directed differentiation protocols for human primed pluripotent stem cells is placental trophoblast and not amnion', Stem Cell Reports, 17(6), pp. 1289-1302. https://doi.org/10.1016/j.stemcr.2022.04.014.

Sokka, J., Lapinsuo, E., Kvist, J., Jalil, S., Yoshihara, M., Weltner, J. et al. (2025) 'Trophoblast stem cell derivation from naive and primed hPSC enables ELF5 functional analysis', Stem Cell Reports, 20(10), article 102637. https://doi.org/10.1016/j.stemcr.2025.102637.

Töhönen, V., Katayama, S., Vesterlund, L., Jouhilahti, E. M., Sheikhi, M., Madissoon, E. et al. (2015) 'Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development', Nature Communications, 6(1), article 8207. https://doi.org/10.1038/ncomms9207.

Takeda, J., Seino, S. and Bell, G. I. (1992) 'Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues', Nucleic Acids Research, 20(17), pp. 4613-4620. https://doi.org/10.1093/nar/20.17.4613.

Tan, J. and Sun, Y. (2025) 'Alternative promoter usage during organ development', PLOS Genetics, 21(3), article e1011635. https://doi.org/10.1371/journal.pgen.1011635.

Theurillat, I., Hendriks, I. A., Cossec, J. C., Andrieux, A., Nielsen, M. L. and Dejean, A. (2020) 'Extensive SUMO Modification of Repressive Chromatin Factors Distinguishes Pluripotent from Somatic Cells', Cell Reports, 32(11), article 108146. https://doi.org/10.1016/j.celrep.2020.108146.

Tietze, E., Barbosa, A. R., Araujo, B., Euclydes, V., Spiegelberg, B., Cho, H. J. et al. (2024) 'Human archetypal pluripotent stem cells differentiate into trophoblast stem cells via endogenous BMP5/7 induction without transitioning through naive state', Scientific Reports, 14(1), article 3291. https://doi.org/10.1038/s41598-024-53381-w.

Torre, D., Francoeur, N. J., Kalma, Y., Gross Carmel, I., Melo, B. S., Deikus, G. et al. (2023) 'Isoform-resolved transcriptome of the human preimplantation embryo', Nature Communications, 14(1), article 6902. https://doi.org/10.1038/s41467-023-42558-y.

Turco, M. Y., Gardner, L., Kay, R. G., Hamilton, R. S., Prater, M., Hollinshead, M. S. et al. (2018) 'Trophoblast organoids as a model for maternal–fetal interactions during human placentation', Nature, 564(7735), pp. 263-267. https://doi.org/10.1038/s41586-018-0753-3.

Varberg, K. M., Dominguez, E. M., Koseva, B., Varberg, J. M., McNally, R. P., Moreno-Irusta, A. et al. (2023) 'Extravillous trophoblast cell lineage development is associated with active remodeling of the chromatin landscape', Nature Communications, 14(1), article 4826. https://doi.org/10.1038/s41467-023-40424-5.

Wang, L., Yi, S., Cui, X., Guo, Z., Wang, M., Kou, X. et al. (2024) 'Chromatin landscape instructs precise transcription factor regulome during embryonic lineage specification', Cell Reports, 43(5), article 114136. https://doi.org/10.1016/j.celrep.2024.114136.

Welch, B. L. (1947) 'The generalization of "Student's" problem when several different population variances are involved', Biometrika, 34(1-2), pp. 28-35. https://doi.org/10.1093/biomet/34.1-2.28.

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4.

Wu, J., Xu, J., Liu, B., Yao, G., Wang, P., Lin, Z. et al. (2018) 'Chromatin analysis in human early development reveals epigenetic transition during ZGA', Nature, 557(7704), pp. 256-260. https://doi.org/10.1038/s41586-018-0080-8.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R. et al. (2013) 'Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells', Nature Structural & Molecular Biology, 20(9), pp. 1131-1139. https://doi.org/10.1038/nsmb.2660.

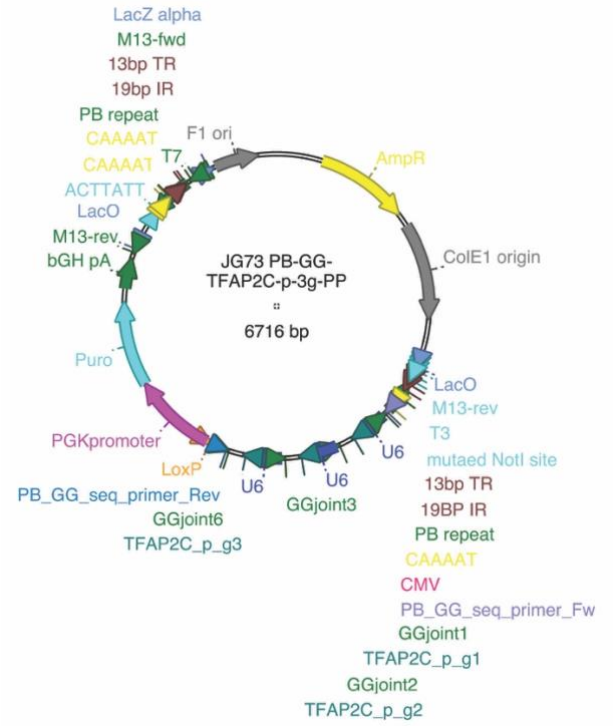
Yu, C., Shen, K., Lin, M., Chen, P., Lin, C., Chang, G. D. and Chen, H. (2002) 'GCMa Regulates the Syncytin-mediated Trophoblastic Fusion', Journal of Biological Chemistry, 277(51), pp. 50062-50068. https://doi.org/10.1074/jbc.m209316200.

Zhu, M., Cornwall-Scoones, J., Wang, P., Handford, C. E., Na, J., Thomson, M. and Zernicka-Goetz, M. (2020) 'Developmental clock and mechanism of de novo polarization of the mouse embryo', Science, 370(6522), article eabd2703. https://doi.org/10.1126/science.abd2703.

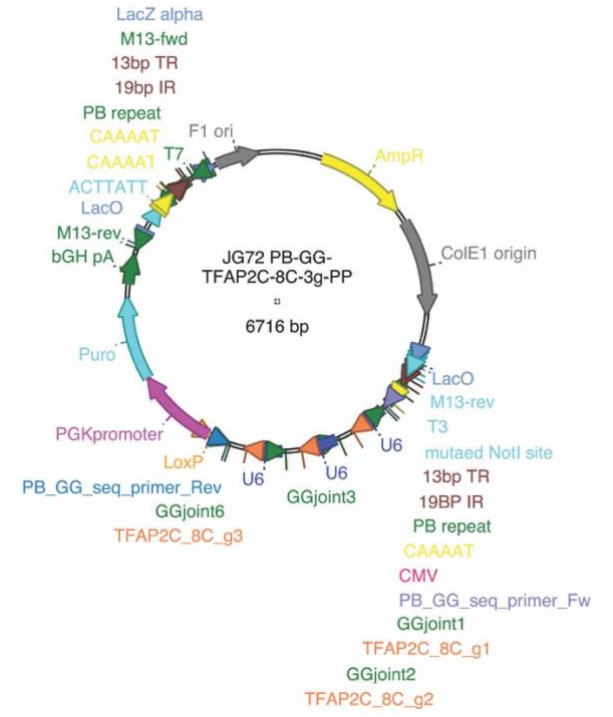
Zhu, M., Meglicki, M., Lamba, A., Wang, P., Royer, C., Turner, K. et al. (2024) 'Tead4 and Tfap2c generate bipotency and a bistable switch in totipotent embryos to promote robust lineage diversification', Nature Structural & Molecular Biology, 31(6), pp. 964-976. https://doi.org/10.1038/s41594-024-01311-9.

9 Appendices

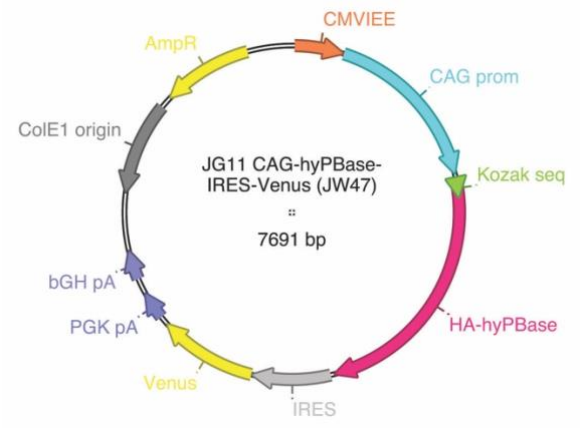
Appendix 1 Plasmid maps



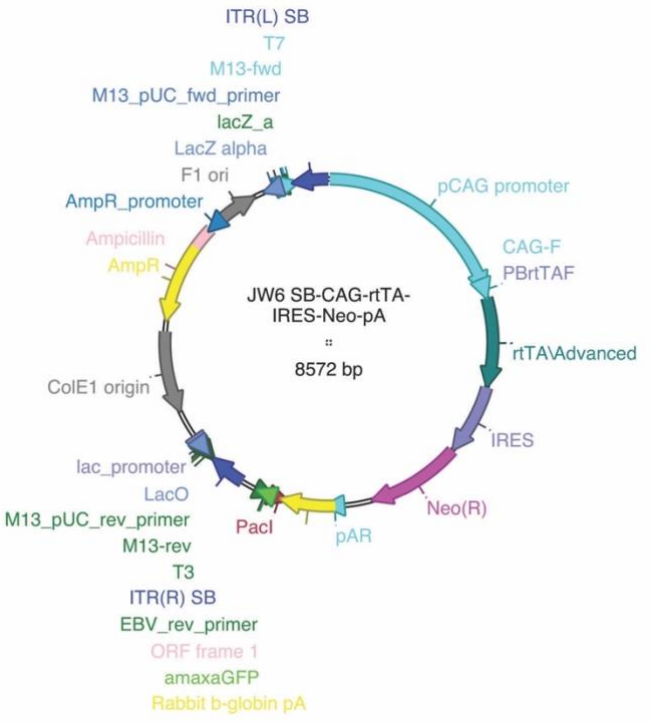
Supplemental Figure 1*. The plasmid map of PB-GG-TFAP2C-P1-3g-PP (PB-GG-TFAP2C-p-3g-PP). The plasmid encodes three gRNAs that target TFAP2C P1 for transcriptional activation.*



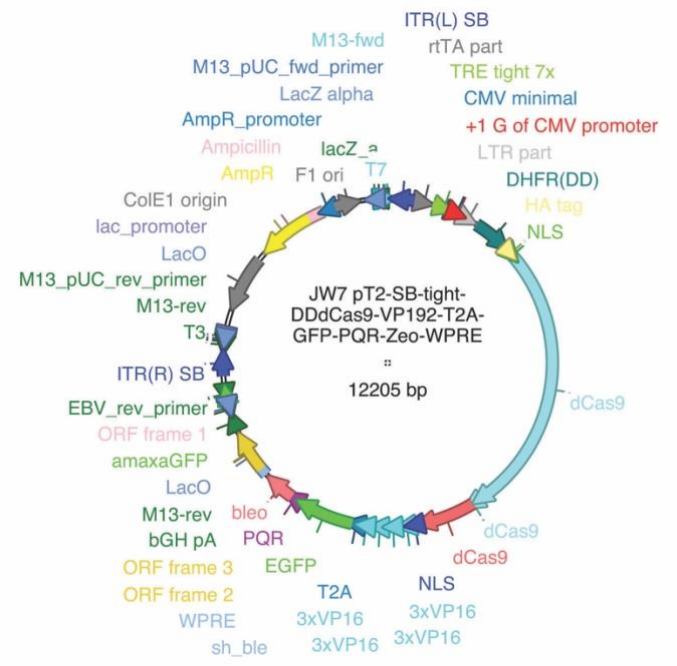
Supplemental Figure 2. *The plasmid map of PB-GG-TFAP2C-P2-3g-PP (PB-GG-TFAP2C-8C-3g-PP). The plasmid codes for three gRNAs, which target TFAP2C P2 for transcriptional activation.*



Supplemental Figure 3*. The plasmid map of CAG-hyPBase-IRES-Venus. The plasmid codes for PiggyBac transposase.*

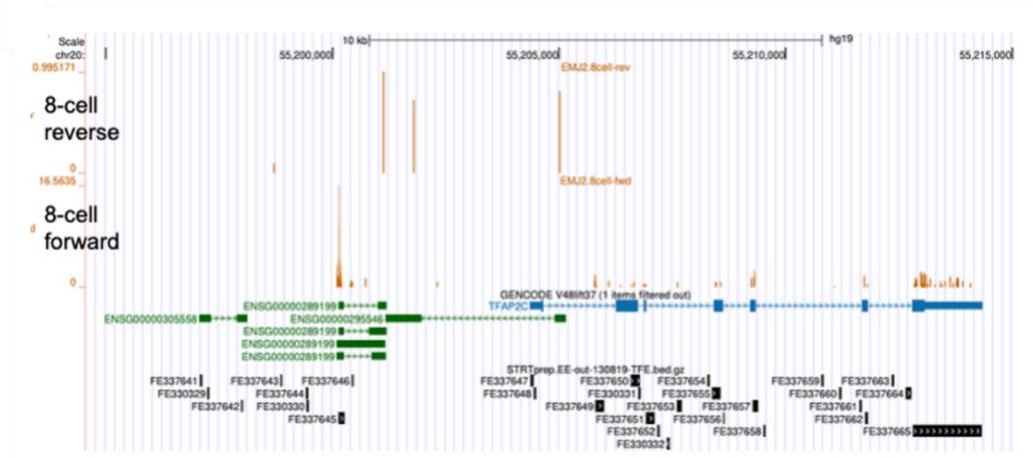


Supplemental Figure 4. *The plasmid map of SB-CAG-rtTA-IRES-Neo-pA. The plasmid codes for a reverse tetracycline-controlled transactivator.*

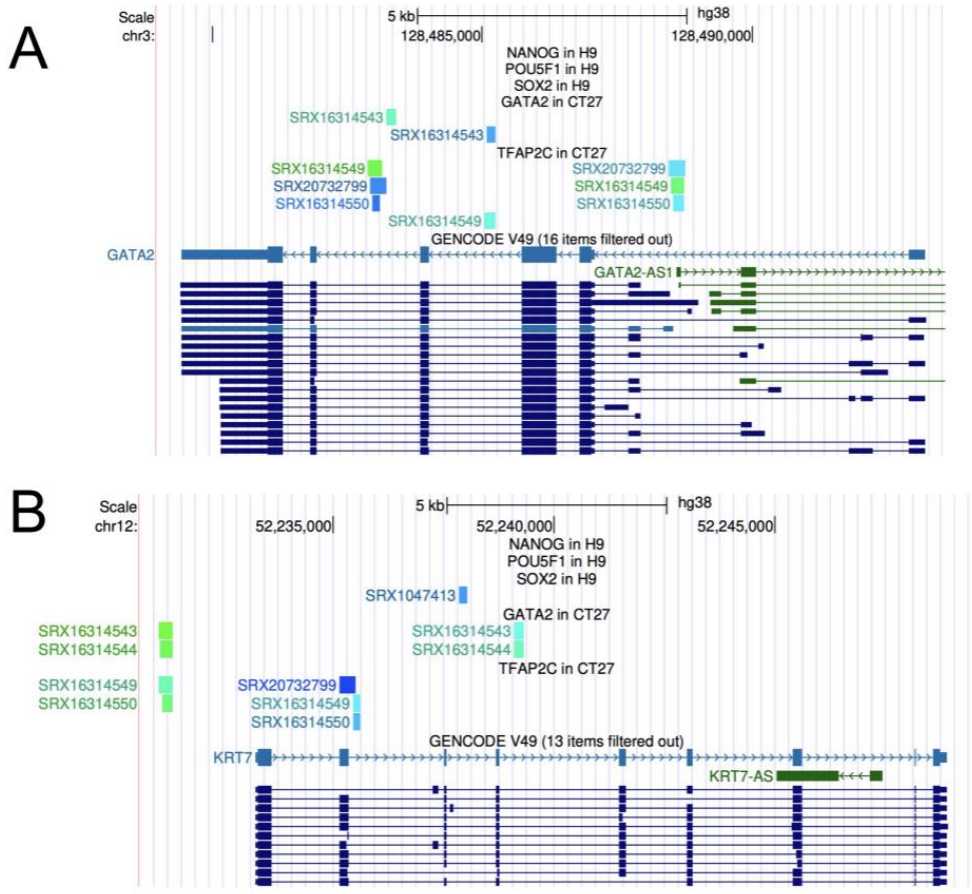


Supplemental Figure 5.*The plasmid map of pT2-SB-tight-DDdCas9-VP192-T2A-GFP-PQR-Zeo-WPRE. The plasmid codes a TRE_tight/minimal CMV-driven dCas9-VP192 expression cassette for CRISPR activation.*

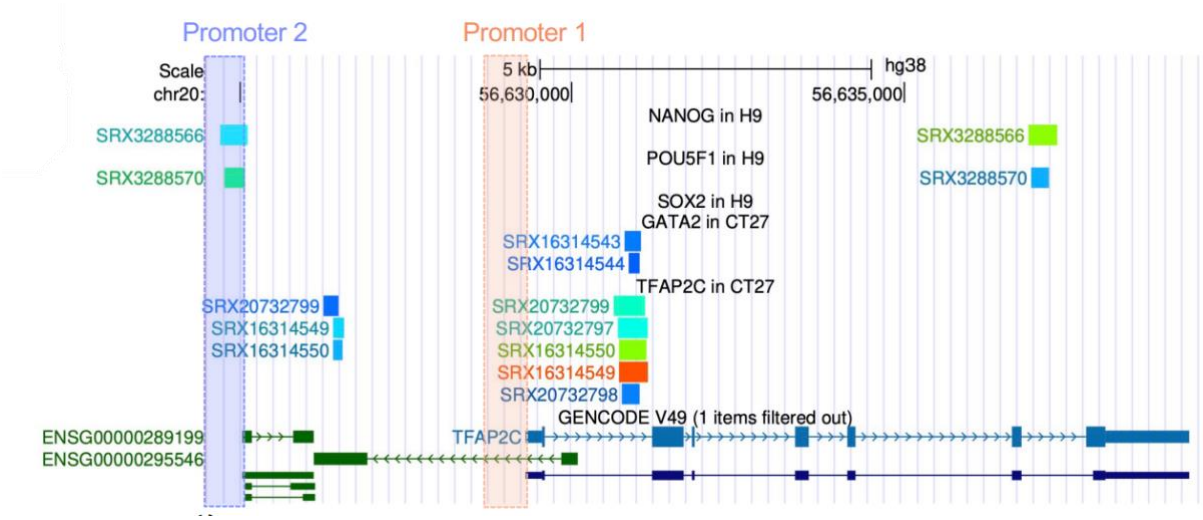
Appendix 2 Additional data from alternative promoter



***Supplemental Figure 6**. Expression of 5' transcript far-ends (TFE) in scSTRT-sequencing data from Jouhilahti et al. (2016).*

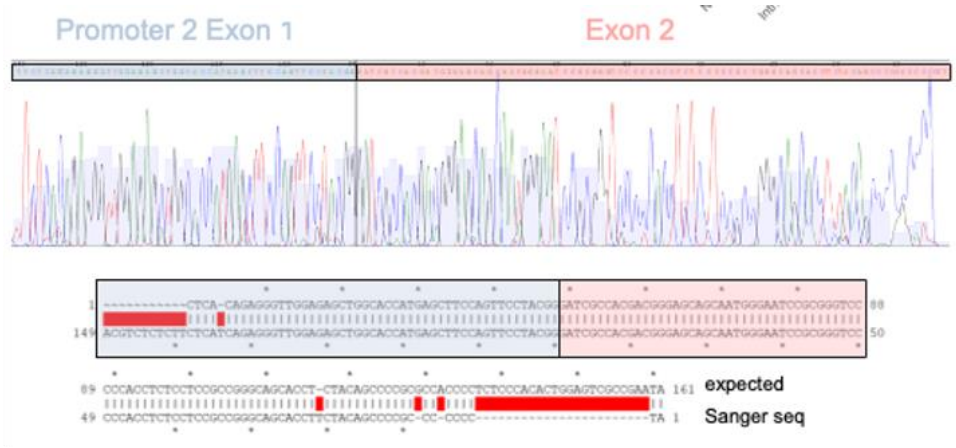


***Supplemental Figure 7**. (A) ChIP-sequencing data of SOX2, NANOG and OCT4 in hESCs and GATA2 and TFAP2C in TSCs at GATA2 locus. (B) ChIP-sequencing data of KRT7 locus.*



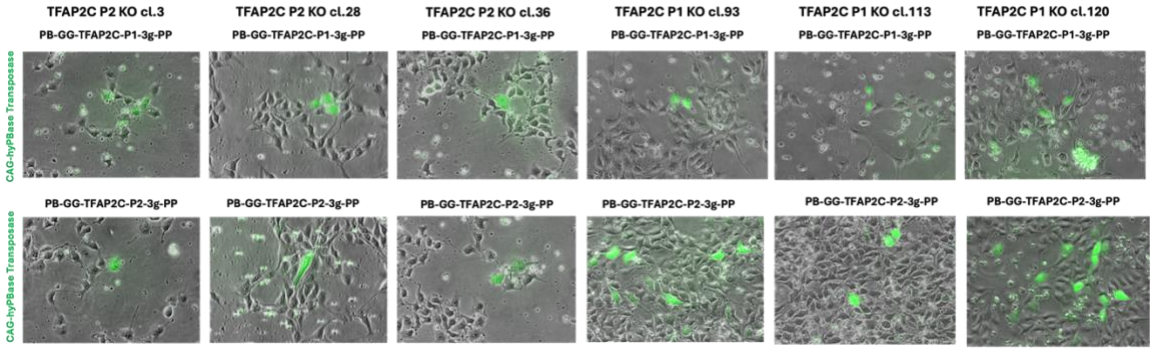
***Supplemental Figure 8**. ChIP-sequencing data of SOX2, NANOG and OCT4 in hESCs and GATA2 and TFAP2C in TSCs at TFAP2C locus with Promoter 2 and Promoter 1 annotation.*

Appendix 3 Sanger sequences



***Supplemental Figure 9**. Sanger sequence over TFAP2C splice site.*

Appendix 4 Immunofluorescence



***Supplemental Figure 10.** Immunofluorescence image of CRISPR activation gRNA plasmid-transfected TFAP2C promoter KO hESCs.*

Appendix 5 Key RT-qPCR test results

Table A1. Key RT-qPCR test results. Across the complete predefined testing family, the CRISPRa condition was compared with the no-guide control of the same genotype and time point. Model p denotes the unadjusted two-sided p-value for the predefined contrast estimated from the blocked linear model. BH denotes the Benjamini-Hochberg-adjusted p-value. Statistical significance was defined as q < 0.05.

Type	Target	Cell and guide	Day	n	log2 effect	FC ratio	Model p	BH q	Sig.	Note
Transcript	P1 transcript	WT; P1 CRISPRa	d3	3/3	8.18	290.4x	1.05e-10	8.38e-10	***	
Transcript	P1 transcript	P2-KO; P1 CRISPRa	d3	3/3	9.83	907.5x	1.96e-14	5.28e-13	***	
Transcript	P1 transcript	P1-KO; P1 CRISPRa	d3	n.t.					n.t.	P1 exon 1 deleted, not tested
Transcript	P1 transcript	WT; P2 CRISPRa	d5	3/3	3.57	11.9x	1.66e-4	5.11e-4	***	
Transcript	P2 transcript	WT; P2 CRISPRa	d3	3/3	9.88	943.0x	2.52e-8	1.44e-7	***	
Transcript	P2 transcript	P1-KO; P2 CRISPRa	d3	3/3	7.98	252.0x	1.18e-9	7.87e-9	***	
Transcript	P2 transcript	P2-KO; P2 CRISPRa	d3	n.t.					n.t.	P2 exon 1 deleted, not tested
Transcript	P2 transcript	WT; P1 CRISPRa	d5	3/3	4.63	24.8x	0.001	0.003	**	
Marker	GATA2	WT; P1 CRISPRa	d7	3/3	1.59	3.0x	0.089	0.106	ns	
Marker	GATA2	P1-KO; P1 CRISPRa	d7	3/3	-1.07	0.5x	0.129	0.142	ns	
Marker	GATA2	P2-KO; P1 CRISPRa	d7	3/3	2.65	6.3x	6.35e-8	1.52e-6	***	
Marker	GATA2	WT; P2 CRISPRa	d7	3/3	3.32	10.0x	0.002	0.005	**	not robust to HC3
Marker	GATA2	P1-KO; P2 CRISPRa	d7	3/3	3.95	15.5x	2.67e-5	1.60e-4	***	
Marker	GATA2	P2-KO; P2 CRISPRa	d7	3/3	0.44	1.4x	0.135	0.142	ns	
Marker	KRT7	WT; P1 CRISPRa	d7	3/3	3.04	8.2x	0.025	0.037	*	not robust to HC3
Marker	KRT7	P1-KO; P1 CRISPRa	d7	3/3	-0.72	0.6x	0.352	0.352	ns	
Marker	KRT7	P2-KO; P1 CRISPRa	d7	3/3	3.86	14.5x	1.17e-5	9.33e-5	***	
Marker	KRT7	WT; P2 CRISPRa	d7	3/3	4.21	18.4x	0.003	0.009	**	not robust to HC3
Marker	KRT7	P1-KO; P2 CRISPRa	d7	3/3	3.66	12.6x	1.60e-4	5.89e-4	***	
Marker	KRT7	P2-KO; P2 CRISPRa	d7	3/3	1.21	2.3x	0.067	0.084	ns	

Table A2. Baseline-adjusted promoter dependence tests. *The baseline-adjusted log2 CRISPRa response was calculated within each biological replicate as (guide – no-guide) at the analysed day minus (guide – no-guide) at day 0.*

Target	Guide	Day	Comparison	n	log2 diff	FC ratio	Welch p	BH q	Sig.
GATA2	P1 CRISPRa	d5	P2-KO vs P1-KO	3/2	3.52	11.5x	0.014	0.075	ns
GATA2	P1 CRISPRa	d5	WT vs P1-KO	3/2	4.09	17.0x	0.041	0.126	ns
GATA2	P1 CRISPRa	d7	P2-KO vs P1-KO	3/3	3.98	15.7x	0.004	0.034	*
GATA2	P1 CRISPRa	d7	WT vs P1-KO	3/3	2.63	6.2x	0.436	0.498	ns
GATA2	P2 CRISPRa	d5	P1-KO vs P2-KO	2/3	0.43	1.4x	0.72	0.72	ns
GATA2	P2 CRISPRa	d5	WT vs P2-KO	3/3	2.04	4.1x	0.168	0.299	ns
GATA2	P2 CRISPRa	d7	P1-KO vs P2-KO	3/3	3.27	9.6x	0.047	0.126	ns
GATA2	P2 CRISPRa	d7	WT vs P2-KO	3/3	3.07	8.4x	0.221	0.354	ns
KRT7	P1 CRISPRa	d5	P2-KO vs P1-KO	3/3	1.48	2.8x	0.312	0.454	ns
KRT7	P1 CRISPRa	d5	WT vs P1-KO	3/3	1.44	2.7x	0.393	0.498	ns
KRT7	P1 CRISPRa	d7	P2-KO vs P1-KO	3/3	4.75	27.0x	0.004	0.034	*
KRT7	P1 CRISPRa	d7	WT vs P1-KO	3/3	3.21	9.2x	0.467	0.498	ns
KRT7	P2 CRISPRa	d5	P1-KO vs P2-KO	3/3	3.05	8.3x	0.028	0.111	ns
KRT7	P2 CRISPRa	d5	WT vs P2-KO	3/3	2.45	5.5x	0.16	0.299	ns
KRT7	P2 CRISPRa	d7	P1-KO vs P2-KO	3/3	2.83	7.1x	0.125	0.287	ns
KRT7	P2 CRISPRa	d7	WT vs P2-KO	3/3	2.89	7.4x	0.426	0.498	ns

Appendix 6 Resource tables

Table A3. Oligonucleotides used in the experiments

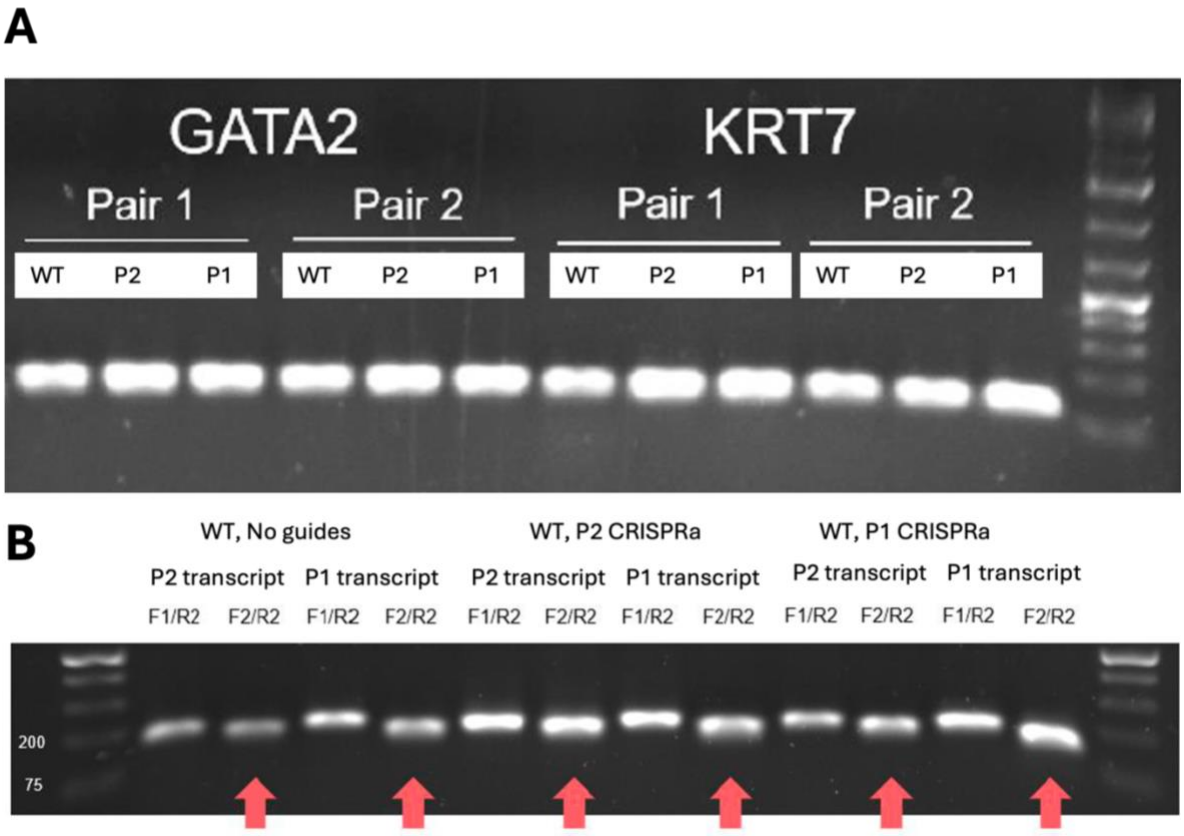
Name	Oligonucleotide	Purpose
TFAP2C P1 transcript Forward primer	AGTACGAAGAGGACTGCGAG	qRT-PCR
TFAP2C P1 transcript Reverse primer	TATTCGGCGACTCCAGTGTG	qRT-PCR
TFAP2C P2 transcript Forward primer	GGCACCATGAGCTTCCAGTT	qRT-PCR
TFAP2C P2 transcript Reverse primer	TATTCGGCGACTCCAGTGTG	qRT-PCR
GATA2 Forward primer	CCACTACCTGTGCAATGCCT	qRT-PCR
GATA2 Reverse primer	AATTTGCACAACAGGTGCCG	qRT-PCR
KRT7 Forward primer	ACATCGAGATCGCCACCTAC	qRT-PCR
KRT7 Reverse primer	ACTGCCACCAGTGGAATTCAT	qRT-PCR
HLA-G Forward primer	TGTCCTTGCAGCTGTAGTCA	qRT-PCR
HLA-G Reverse primer	AAGGGACTTGCCACTCAGTC	qRT-PCR
hCGB Forward primer	GCTGGACCAGTGAGAGGAGAG	qRT-PCR
hCGB Reverse primer	CAAGGCCACCAGGAGGTGA	qRT-PCR
PPIG Forward primer	TCTTGTCAATGGCCAACAGA	qRT-PCR
PPIG Reverse primer	GCCCATCTAAATGAGGAGTT	qRT-PCR
P1 gRNA1 oligo sequence, guide target sequence bolded	GTGGAAAGGACGAAACACCG **ACGCTGTGAGGGACGAAGCG** GTTTTAGAGCTAGAAATAG	CRISPRa
P1 gRNA2 oligo sequence, guide target sequence bolded	GTGGAAAGGACGAAACACCG **AATCTCAGGTGTGTTTACGG** GTTTTAGAGCTAGAAATAG	CRISPRa
P1 gRNA3 oligo sequence, guide target sequence bolded	GTGGAAAGGACGAAACACCG **GGGGAGCGACTCAACAACCG** GTTTTAGAGCTAGAAATAG	CRISPRa
P2 gRNA1 oligo sequence, guide target sequence bolded	GTGGAAAGGACGAAACACCG **CAGAGTTTATAGTTTCCAAG** GTTTTAGAGCTAGAAATAG	CRISPRa
P2 gRNA2 oligo sequence, guide target sequence bolded	GTGGAAAGGACGAAACACCG **ATCCCTTGCCAGTAAGAGGT** GTTTTAGAGCTAGAAATAG	CRISPRa
P2 gRNA3 oligo sequence, guide target sequence bolded	GTGGAAAGGACGAAACACCG **GTTCAGCTGAAATTGAACCA** GTTTTAGAGCTAGAAATAG	CRISPRa
TFAP2C P2 KO primer F1	GAAGCTGTTTGGGTTGCCTG	KO PCR
TFAP2C P2 KO primer F2	CCCGGAAACCTGTGATGTGA	KO PCR
TFAP2C P2 KO primer F3	CCTGCCAGTCTTCCAATCGC	KO PCR
TFAP2C P2 KO primer R1	GCCGCATGCTCTTCGTTTAG	KO PCR
TFAP2C P2 KO primer R2	CTGAAAGCCACAAAGGTGCC	KO PCR
TFAP2C P2 KO primer R3	CTTCGCAGCGAGTCCTCC	KO PCR
TFAP2C P1 KO primer R5	TTGGAGGAACGAGTGAAGGC	KO PCR
TFAP2C P1 KO primer F5	TTCCGAGAAAACGGGAAGCA	KO PCR
TFAP2C P1 KO primer F4	CCTTAAGCCTCAGTGGGGTG	KO PCR
TFAP2C P1 KO primer R4	TTAGGAAAGCGCCCGGTAAG	KO PCR
KO gRNA4 (P1) oligo sequence, guide target sequence bolded	TAATACGACTCACTATAG **ATTATCATAACTCCTCCCCG** GTTTTAGAGC TAGAA	KO
KO gRNA5 (P1) oligo sequence, guide target sequence bolded	TAATACGACTCACTATAG **TTTCCAGGCCTCGTTCCCGA** GTTTTAGAGC TAGAA	KO
KO gRNA1 (P2) oligo sequence, guide target sequence bolded	TAATACGACTCACTATAG **GTAGAAAAAGAGTTCCCCGA** GTTTTAGAGC TAGAA	KO
KO gRNA2 (P2) oligo sequence, guide target sequence bolded	TAATACGACTCACTATAG **TGAGAAAAGAGGTGTCTGTG** GTTTTAGAGC TAGAA	KO
KO gRNA3 (P2) oligo sequence, guide target sequence bolded	TAATACGACTCACTATAG **TGGCGACCCAGAACACGGAG** GTTTTAGAGC TAGAA	KO

Table A3. Key resources table

Reagent or resource	Source	Identifier
Antibodies		
Rabbit polyclonal anti-GATA2 (1:400)	Sigma-Aldrich	Cat# HPA005633
Rabbit polyclonal anti-hCG beta (1:400)	Thermo Fisher Scientific	Cat# PA5-58598
Mouse monoclonal anti-HLA-G (1:400)	Santa Cruz	Cat# sc-21799
Mouse monoclonal anti-KRT7 (1:800)	Thermo Fisher Scientific	Cat# MA1-06316
Goat polyclonal anti-TFAP2C (1:1000)	R&D Systems	Cat# AF5059
Mouse monoclonal anti-αSMA (1:1000)	Sigma-Aldrich	Cat# A2547
Mouse monoclonal anti-βIII-tubulin (1:1000)	Thermo Fisher Scientific	Cat# 14-4510-82
Rabbit monoclonal anti-SOX17 (1:400)	Cell Signaling	Cat# 81778
Donkey anti-Goat IgG (H+L) Alexa Fluor 488 (1:500)	Thermo Fisher Scientific	Cat# A11055
Donkey anti-Rabbit IgG (H+L) Alexa Fluor 488 (1:500)	Thermo Fisher Scientific	Cat# A21206
Donkey anti-Mouse IgG (H+L) Alexa Fluor 488 (1:500)	Thermo Fisher Scientific	Cat# A21202
Donkey anti-Rabbit IgG (H+L) Alexa Fluor 594 (1:500)	Thermo Fisher Scientific	Cat# A21207
Donkey anti-Mouse IgG (H+L) Alexa Fluor 594 (1:500)	Thermo Fisher Scientific	Cat# A21203
Chemicals, peptides, and recombinant proteins		
Gibco Minimum Essential Medium	Thermo Fisher Scientific	Cat# 11090081
Biolaminin 521	BioLamina	Cat# LN521
Gibco Essential 8	Thermo Fisher Scientific	Cat# A1517001
Gibco Versene Solution (EDTA)	Thermo Fisher Scientific	Cat# 15040066
ROCK inhibitor, Y-27632	Cayman	Cat# 10005583
Puromycin dihydrochloride	Thermo Fisher Scientific	Cat# J67236.XF
Doxycycline	Fisher Bioreagents	Cat# BP2653-1
Trimethoprim	Fisher Scientific	Cat# 15345602
Gibco Essential 6	Thermo Fisher Scientific	Cat# A1516401
CHIR99021	Stemgent	Cat# 04-0004-10
A-83-01	Sigma	Cat# SML0788-5MG
SB431542	Cayman Chemicals	Cat# 13031
Sodium Butyrate	Sigma	Cat# B5887-1G
TrueCut™ Cas9 Protein v2	Thermo Fisher Scientific	Cat# A36498
Alt-R™ Cas9 Electroporation Enhancer, 10 nmol	Integrated DNA Technologies	Cat# 1075916
Ambion™ Nuclease-Free Water	Life Technologies	Cat# AM9937
Triton X-100	Sigma-Aldrich	Cat# T9284
Tween20	Sigma-Aldrich	Cat# P1379
Hoechst 33342	Thermo Fisher Scientific	Cat# 62249
RT buffer	Promega	Cat# M531A
dNTP mix	BioNordika	Cat# BN-1006-08
Oligo dT primer	Thermo Fisher Scientific	Cat# SO132
random hexamers	Promega	Cat# C118A
RNAse inhibitor	Thermo Fisher Scientific	Cat# EO0381
M-MLVR reverse transcriptase	Promega	Cat# M170A
5x buffer with 7.5 mM MgCl2	Thermo Fisher Scientific	Cat# F524
20x EvaGreen	Biotium	Cat# 31000
Phire Hot Start II DNA polymerase	Thermo Fisher Scientific	Cat# F122L
Commercial assays		
NucleoSpin RNA Plus kit	Macherey-Nagel	Cat# 740984
Neon™ Transfection System 100 µl Kit	Thermo Fisher Scientific	Cat# MPK10096
Neon™ Transfection System 10 µl Kit	Thermo Fisher Scientific	Cat# MPK1096
GeneArt™ Precision gRNA Synthesis Kit	Invitrogen	Cat# A29377
Experimental models: Cell lines (human)		
DDdCas9VP192-TetON-H9 (H9 SBact)	Gawriyski et al., 2023	N/A
H9 SBact TFAP2C P1 KO cl. 93	This thesis	N/A
H9 SBact TFAP2C P1 KO cl. 113	This thesis	N/A
H9 SBact TFAP2C P1 KO cl. 120	This thesis	N/A
H9 SBact TFAP2C P2 KO cl. 3	This thesis	N/A
H9 SBact TFAP2C P2 KO cl. 28	This thesis	N/A
H9 SBact TFAP2C P2 KO cl. 36	This thesis	N/A

Reagent or resource Software and algorithms	Source	Identifier
Fiji StarDist 2D R (Version 4.4) ggplot2 (Version 3.5.2) Benchling UCSC LiftOver	Schindelin et al., 2012 Schmidt et al., 2018 R Core Team (2024) Wickham, 2016 Benchling (2026) UCSC (2026)	https://imagej.net/software/fiji/ https://imagej.net/plugins/stardist https://cran.r-project.org https://ggplot2.tidyverse.org/ https://benchling.com https://genome.ucsc.edu/cgi-bin/hgLiftOver?hgsid=4008109327_QS3ChFV1J60HrCmYAPJe1rsUSdGY

Appendix 7 RT-qPCR Primer validation



Supplemental Figure 11. Validation of RT-qPCR primers. (A) GATA2 and KRT7 primer validation. *WT=TFAP2C wt cells, no guides, P2 = wt cells, P2 CRISPRa, P1 = wt cells, P1 CRISPRa. Primer pair 2 were used for both genes.* ***(B) TFAP2C P1 and P2-derived transcript primer validation****. Forward primers were isoform specific. Reverse primers were the same. F2/R2 primers were used for both transcripts.*