UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea magistrale in Molecular Biology

TESI DI LAUREA

Exploring the timing and dynamics of X chromosome inactivation during the specification of human trophectoderm cells

Relatrice: Prof.ssa Chiara Rampazzo

Dipartimento di Biologia

Laureanda: Romina Facchinello

ANNO ACCADEMICO 2022-2023

TABLE OF CONTENTS

SUMMARY.....3

GLOSSARY OF TERMS.....4

1. INTRODUCTION.....7

1.1 Historical background: the discovery of the Barr body and Mary Lyon's hypothesis.....7

1.2 X chromosome inactivation across the evolution of mammals.....8

1.3 Dynamics of X chromosome inactivation during mouse development.....9

1.4 Interspecies variability: the study case of *Macaca fascicularis*.....15

1.5 The state of the art for XCI in human.....16

1.6 Human Trophectoderm Cells and Trophoblast Stem Cells as a model for early extraembryonic development.....19

1.7 Presentation of the experimental plan: aims and preliminary results.....21

2. MATERIAL AND METHODS.....23

2.1 Cell culture.....23

2.2 RNA-FISH probe synthesis by nick translation.....26

2.3 RNA-Fluorescence *In Situ* Hybridization (RNA-FISH).....26

2.4 Immunofluorescence.....27

2.5 RT-qPCR.....28

2.6 Fluorescence-Activated Cells Sorting (FACS).....28

2.7 Statistical analysis.....28

3. RESULTS.....30

3.1 Optimization of naïve hESC conversion to trophectoderm-like identity30

2.1 Induction of *XIST* expression via doxycycline treatment is functional in trophectoderm cells.....32

2.2 Characterization of the *XIST* knockdown phenotype on TE-like converted cells.....33

4. DISCUSSION.....39

BIBLIOGRAPHY.....44

SUMMARY

In placental mammals, X-linked gene expression is balanced between females (XX) and males (XY) through a process termed "X Chromosome Inactivation" (XCI). This event leads to the almost complete transcriptional silencing of one of the two X chromosomes during early development of female embryos. The master regulator of XCI initiation is *Xist* ("X-Inactive Specific Transcript"), a long non-coding RNA able to trigger in *cis* a cascade of epigenetic modifications that ultimately leads to the X chromosome-wide transcriptional repression.

Despite XCI being essential, both mouse and primate non-human embryos showed crucial divergences regarding when and how XCI is set in the embryonic and extraembryonic lineages. As for humans, the timing and mechanisms involved in triggering XCI remain elusive. By using human Embryonic Stem Cells (hESCs) in a naïve-like state of pluripotency, the present study aimed to characterize *XIST* role for XCI and its requirement during the specification of early trophectoderm (TE) *ex vivo*. We could determine that XCI is initiated during TE specification and that *XIST* is essential for the process. Moreover, disruption of *XIST* expression does not seem to dramatically impair TE specification. This suggests that XCI is not strictly required for the acquisition of TE identity.

GLOSSARY OF TERMS

Gene	Description (GeneCards®)
ATRX	ATRX chromatin remodeler
CDX2	Caudal Type Homeobox 2
CTCF	CCCTC-Binding Factor
ENPEP	Glutamyl Aminopeptidase (Aminopeptidase A)
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GATA2	GATA Binding Protein 2
GATA3	GATA Binding Protein 3
HUWE1	HECT, UBA And WWE Domain Containing E3 Ubiquitin Protein Ligase 1
MECP2	Methyl-CpG Binding Protein 2
MEK	Mitogen-Activated Protein Kinase Kinase 1
NANOG	Homeobox Transcription Factor Nanog
NODAL	Nodal Growth Differentiation Factor
NR2F2	Nuclear Receptor Subfamily 2 Group F Member 2
POU5F1 (OCT4)	POU Class 5 Homeobox 1
POLA1	DNA Polymerase Alpha 1, Catalytic Subunit
RBM15	RNA Binding Motif Protein 15
ROCK	Rho Associated Coiled-Coil Containing Protein Kinase

SOX2	SRY-Box Transcription Factor 2
SPEN	Spen Family Transcriptional Repressor
TFAP2A	Transcription Factor AP-2 Alpha
TFAP2C	Transcription Factor AP-2 Gamma
VIM	Vimentin
WTAP	WT1 Associated Protein
XACT	X Active Specific Transcript
XIST	X Inactive Specific Transcript

Protein domains and protein complexes	**Description**
KRAB	Krüppel Associated Box domain
PRC1, PRC2	Polycomb Repressive Complex 1, 2

Histone modifications	**Description**
H2AK119ub	Mono-ubiquitination of lysine 119, histone H2A
H3K27me3	Trimethylation of lysine 27, histone H3
H3K4me3	Trimethylation of lysine 4, histone H3
H3K9 acetylation	Acetylation of lysine 9, histone H3
H3K9me3	Trimethylation of lysine 9, histone H3

Cell denominations	Description
ESC	Embryonic Stem Cell
ICM	Inner Cell Mass
TE	Trophectoderm
TSC	Trophoblast Stem Cell

X chromosome-related abbreviations	Description
XCI	X Chromosome Inactivation
XCU	X Chromosome Upregulation
XIC	X chromosome Inactivation Center

Other	Description
BSA	Bovine Serum Albumin
Cas9	CRISPR-associated endonuclease 9
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi	CRISPR Interference
CSK buffer	Cytoskeletal buffer
dNTP	Deoxynucleoside triphosphate
PBS buffer	Phosphate-Buffered Saline buffer
PFA	Paraformaldehyde
rtTA	reverse-tetracycline transactivator
SSC buffer	Saline-Sodium Citrate buffer
VRC	Ribonucleoside Vanadyl Complex

1. INTRODUCTION

1.1 Historical background: the discovery of the Barr body and Mary Lyon's hypothesis

While studying the morphology of cat-derived motor neurons, physician Murray L. Barr and his student Ewart G. Bertram came across a female-specific nuclear trait: a condensed body usually seen close to the nucleolus and typically absent in male corresponding neurons (***fig. 1***). This sex-related dimorphism, together with the consideration that sexual chromosomes were frequently found near the nucleolus, let them hypothesize that the nature of such a nuclear body consisted indeed in one X chromosome[1].

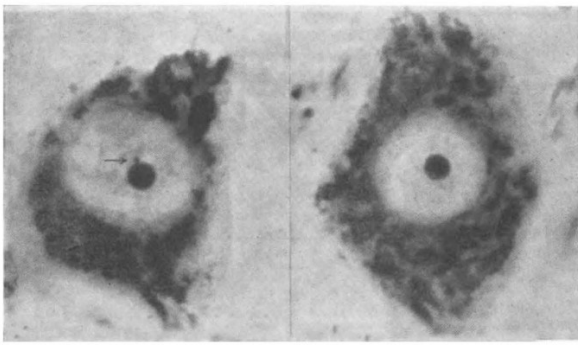

Figure 1*: Motor neurons from the hypoglossal nucleus of mature female (left) and male (right) cats stained with cresyl violet. The observed female-specific nuclear body is indicated with an arrow (Barr et al., 1949).*

The "Barr body", as it was called, was later observed in other cell types, including mouse somatic cells. This cytologic feature was brilliantly associated with X-linked gene silencing by Mary Lyon in 1961[2]: at the time, she was studying the transmission of X-linked coat color traits, and she noticed a peculiar mosaic phenotype in heterozygous female mice. Knowing that XO females could successfully reach adulthood, she proposed the mosaicism to be the direct consequence of the random inactivation of one out of two X chromosomes. This process would have occurred early during development, bringing to the formation of the heteropycnotic Barr body, and explaining the coat patches as groups of cells descending from a common ancestor in which one or the other X-linked allele had been inactivated.

1.2 X chromosome inactivation across the evolution of mammals

Since then, a body of evidence has been collected showing coherence with Mary Lyon's conclusions. The organized compaction and silencing of one X chromosome in females were demonstrated to be a recurrent feature during the embryonic development of both placental and marsupial mammals[3]. However, the epigenetic program at the basis of XCI establishment greatly diverged throughout evolution. In a variety of marsupial species, a nonrandom imprinted form of epigenetic silencing takes place, which selectively targets the paternally inherited X chromosome. What is more, X inactivation appears to be unstable and incomplete: X-linked genes are not silenced to the same extent in all tissues and are frequently reactivated[3]. In placental mammals, with some remarkable exceptions, one of the two Xs is randomly and more stably inactivated independently of the parental origin. Importantly, in both subclasses the process appears to be directed by evolutionary unrelated long non-coding RNAs: *Rsx* in marsupials (RNA on the Silent X) and *Xist* in placentates (X Inactive Specific Transcript), which tend to propagate in *cis* over the future inactive X chromosome during early embryonic development. Although the underlying mechanisms vary according to the subdivision and even the species considered, they nevertheless provide a solution to a common issue: gene dosage imbalance between sexes.

Among mammals, sex is genetically determined by a pair of heterologous chromosomes, termed X and Y: the female sex is homogametic while the male sex is heterogametic. The XX-XY system presumably evolved from ordinary autosomes, with the proto-Y chromosome acquiring by chance the competency to guide male sex development and fertility[4]. The frequency of crossing over events between the proto-Y and proto-X chromosomes progressively decreased with further Y chromosomal rearrangements and this likely contributed to their divergence. In this scenario, the Y chromosome became relatively small (60 million bp in *Homo sapiens*), enriched in repetitive DNA and highly specialized: it only contains a limited number of genes, mostly required to sustain male sex determination and spermatogenesis. The X chromosome, on the other hand, is provided with a higher number of genes (approximately a thousand) involved in embryonic development, female-specific sexual maturation, immunity and other processes[5].

Most genes on the X chromosome do not have an homologous counterpart on the Y chromosome, which means that by default females are endowed with twice as much allelic content compared to males. This asset would imply *a priori* approximately double gene product in the female sex, which has been proven to dramatically impair embryonic development[6]. A key control system likely evolved together with sex chromosomes to equalize X-linked gene expression of female embryonic cells at the transcriptional level.

Of note, XCI is not the only regulatory phenomenon known to affect X chromosome-wide gene expression. Both males and females undergo a transcriptional upregulation of X-linked genes early during development, which has been interpreted as a mechanism aimed at balancing the levels of X-linked and autosomal gene expression. Hence, in females, proper gene expression balance would be the result of a dynamic crosstalk between XCU (X chromosome upregulation) and XCI[7].

1.3 Dynamics of X chromosome inactivation during mouse development

Here are presented the key common steps in the early placental mammalian development[8], first described in the mouse. Due to its relatively high genetic manipulability, prolificacy and low-cost maintenance compared to other mammals, *Mus musculus* has dominated the scientific field as a model for mammalian-specific processes. Nonetheless, increasing evidence highlighted substantial species-specific characteristics that urged the employment of additional tools when referring to another species, as *Homo sapiens.* For instance, mouse embryonic development progresses faster compared to other mammals, including primates: the timeline in ***fig. 2*** is therefore specific to the mouse model.

The zygote first undergoes a few rounds of asynchronous cell divisions: at this point, all cells are totipotent. After a few cleavage events, the structure compacts through intercellular junctions and undergoes further cell divisions. This sets the first event of fate divergency between an external sealed layer of polarized cells and an inner population of adherent cells, still nonpolarized.

These two cell districts can be morphologically distinguished at the early blastocyst stage (32 cells), immediately before implantation (***fig. 2***). The external layer, or trophectoderm, is committed to specialize into Trophoblast Stem Cells (TSCs):

these will differentiate constituting the chorion, that is the fetal contribution to the placenta. The innermost cells, on the other hand, further compact into the so-called Inner Cell Mass (ICM), which remains partially separated from the trophectoderm by a liquid-filled cavity, the blastocoel.

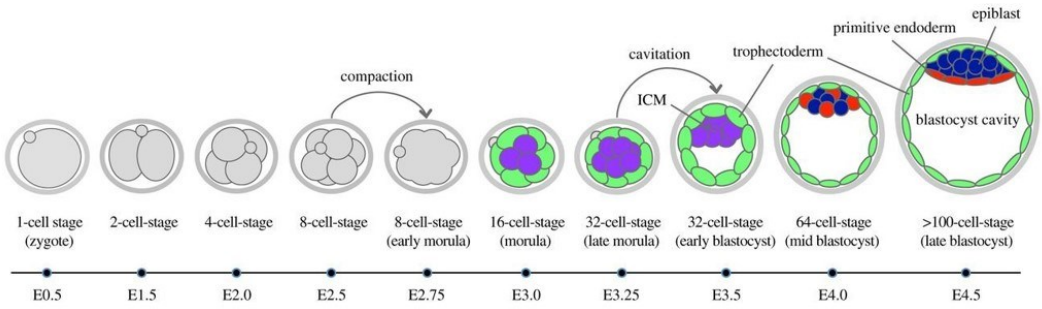

***Figure 2**: Timeline expressed in embryonic days (E-) for the main events occurring before implantation during mouse embryonic development (Mihajlović et al., 2017).*

Importantly, trophectoderm formation is the first specification event taking place during mammal development and is fundamentally required for implantation. In mouse, the trophectoderm-ICM dichotomy is underlined by the mutually exclusive expression of *Cdx2* (trophectoderm marker) and *Pou5f1* (ICM marker), the origin of which is still discussed[9]. A second important fate divergence occurs in the ICM before the gastrulation phase: a polarized cell layer, the primitive endoderm (or hypoblast), forms facing the blastocoel. This tissue will specialize into the yolk sac, while only the innermost cell population, or epiblast, will provide the embryonic germ layers.

These early stages of development have been explored using both *in vivo* and *ex vivo* models. Mouse embryos can be cultured until the blastocyst stage, within the pre-implantation temporal window. In addition, since 1981 mouse Embryonic Stem Cells (mESCs) have been isolated from the ICM and maintained *in vitro*[10,11], thus offering another opportunity to characterize the profile of pluripotent stem cells.

The pluripotency state is a transient non-differentiated condition established by a set of key transcriptional factors (pluripotency factors), the most important being POU5F1, NANOG and SOX2, which are functionally conserved in *H. sapiens*. If expressed in a somatic cell, they force a shift of the proliferative, metabolic and transcriptional profile into a less differentiated state. This was observed in a number of tumoral cases[12] and has been also successfully exploited for the controlled *in vitro* reprogramming of differentiated cells towards pluripotency[13].

Pluripotency factors are entangled in a transcriptional circuit that (a) promotes self-renewal, (b) prevents differentiation and (c) maintains their expression. Their activity is ultimately conveyed by epigenetic factors, which deeply affect the chromatin state[14]. In this scenario, XCI is crucially linked to the early events of fate determination. In pluripotent female stem cells two active X chromosomes can be observed, a condition that is verified both *in vivo* and *ex vivo*[15]. With differentiation, the transcriptional profile specializes through the formation of cell type-specific euchromatic and heterochromatic domains, the latter including one X chromosome. The inability to successfully achieve XCI was shown to severely impair mouse blastocyst formation and tendentially block embryogenesis soon after implantation[6]. Consistently, mESCs are unable to properly differentiate in the absence of XCI[6], further supporting evidence for XCI being coupled to differentiation.

More in detail, the progression of XCI can be subdivided into three different stages: (1) initiation, (2) establishment and (3) maintenance.

1) *Initiation*

The initiation phase includes a mechanism of counting and choosing the X chromosome(s) to be inactivated. X-linked gene dosage regulation also depends on the number of chromosomal arrays. In a diploid female cell (two chromosomal arrays, one paternally and one maternally inherited), only one X chromosome is inactivated. In case of X chromosome aneuploidy (n ≠ 2 Xs), all but one will undergo XCI. Tetraploidy (four chromosomal arrays) leads to the inactivation of two rather than one X chromosome. The counting implies a need for balance with the autosomal gene dosage, and the underlying mechanisms are still under discussion[16].

Interestingly, early mouse embryos undergo two subsequent waves of XCI: a first imprinted event occurs soon after the transcriptional activation of the zygotic genome (2 cell stage). In all blastomeres, the paternally inherited X chromosome is preferentially silenced. While this state is maintained in the extraembryonic lineage[15], after implantation cells of the ICM reverse XCI only for it to be subsequently re-established in a random fashion. This implies two important aspects: (1) XCI is not necessarily stable in physiological conditions; (2) the

embryonic and extra-embryonic lineages fundamentally diverge even regarding how and when XCI is initiated.

Nevertheless, both forms of XCI are triggered by a common master regulator, a long non-coding RNA named *Xist* according to its preferential accumulation on the inactive X[6,17]. *Xist* locus resides within the "X chromosome Inactivation Center" (Xic), together with genes involved in the regulation of its expression. Xic, identified using mESCs[18], is defined as the minimal regulative region required for XCI initiation. This genomic region spans ~ 800 kb spatially organized in two topologically associated domains (TADs)[15,17]. These 3D nuclear domains are also functionally distinct, as they tendentially host loci involved either in the positive or negative regulation of XCI. The first category comprises *Xist* lncRNA and positive regulators of its expression, most notably *Jpx* and *Ftx* non-coding loci. Among the negative modulators of *Xist* and XCI, *Tsix* lncRNA is the most remarkable in mouse: this antisense RNA, transcribed from *Xist* complementary strand, prevents *Xist* upregulation before XCI initiation. Indeed, in undifferentiated mESCs *Xist* is not expressed, its downregulation being assured by a collection of pluripotency factors (POU5F1, SOX2, NANOG) and long non-coding RNAs transcribed from the Xic region[6]. While exiting from pluripotency under differentiation inputs, the repressive grip on *Xist* expression loosens. Importantly, only one X chromosome will significantly upregulate *Xist*, the other maintaining its repression through *Tsix*[18]. How this differential distribution is achieved is still under study.

2) *Establishment*

Xist is a multitasking molecule endowed with functionally distinct RNA domains partially overlapping with blocks of tandem repeats (A-F, ***fig. 3***). Once expressed, it tends to coat the X chromosome in *cis* and recruit downstream effectors of XCI. Importantly, *Xist* has been demonstrated to be necessary and sufficient to trigger XCI[17], although in a context-specific role: outside a critical temporal window during early ESC differentiation, it is no longer able to induce XCI[6].

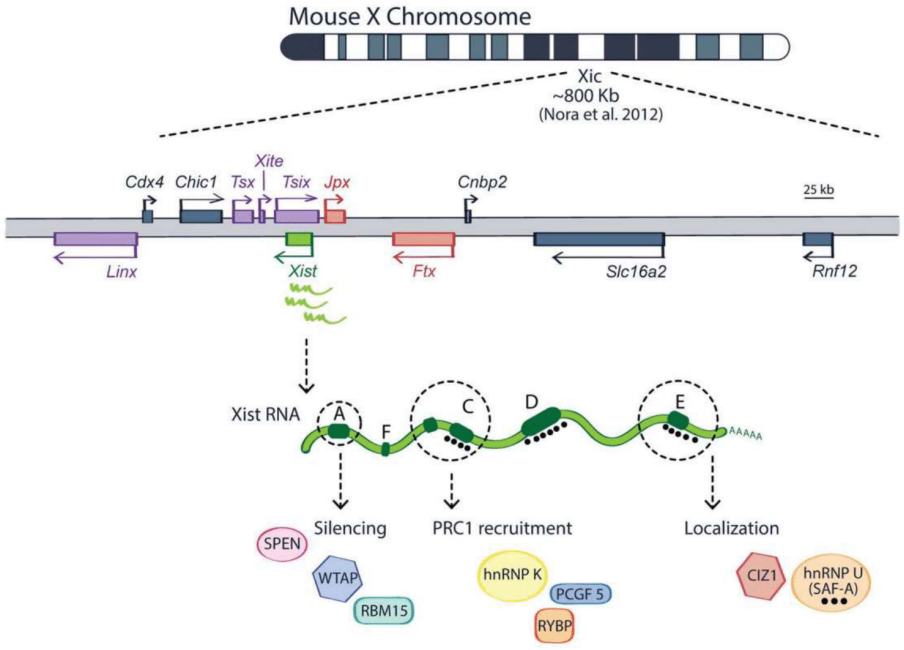

Figure 3: *On top, simplified organization of the mouse Xic region. Positive regulators of Xist expression are indicated in red, while negative regulators are highlighted in violet. Xist domains are named with alphabetical letters and associated to their identified functions and interactors (Loda et al., 2019).*

The spreading of *Xist* along the X chromosome allows for a chromosome-wide rewriting of the chromatin landscape. The first most important events of XCI consist in the local loss of RNA polymerase II, the erasure of euchromatin-associated histone marks (H3K4me3, H3K9 acetylation) and the subsequent deposition of repressive histone modifications (H2AK119ub, H3K27me3). These initial events are reinforced by the progressive and hierarchical addition of layers of repressive modifications, including the deposition of macroH2A histone variant and massive DNA methylation[6]. Proteomic studies highlighted several *Xist* interaction partners directly or indirectly involved in chromatin compaction[17]: among the most relevant are PRC1, responsible for the mono-ubiquitination of H2AK119; SPEN, proposed to indirectly promote *Xist* upregulation[19] and histone deacetylation; RBM15 and WTAP, involved in an RNA methylation pathway which may affect XCI at the post-transcriptional level.

Importantly, with XCI other important changes are imposed: the replication timing of the inactive X chromosome shifts towards the late S-phase, while its nuclear territory tends to occupy distinct compartments at the nuclear membrane or close to the nucleolus[6,20].

As a matter of fact, the generalized transcriptional repression does not involve the entirety of the X-linked genes: loops can emerge from X-inactive chromosome regions and still engage with transcription factories. These sporadic euchromatic traits are maintained by architectural proteins, such as CTCF[20]. X-linked loci being still transcribed are called "escaping genes": they may have an homologous counterpart on the Y chromosome, or play important roles in female-specific developmental aspects[4,17]. Of note, the number of escapees and their expression level greatly vary according to the species and the developmental stage[6].

3) *Maintenance*

Imprinted and random XCI were found to differ in the kinetics and the extent of repressive mark distribution (H3 deacetylation, H3K4 demethylation, H3K27me3, macroH2A loading, DNA methylation), thus generating two different signatures on the inactive X[6]. Interestingly, *in vivo* analyses and *in vitro* studies conducted on trophectoderm and hypoblast derivatives showed that imprinted XCI seems to be more labile and plastic, allowing for a higher number of escapees. This may suggest imprinted XCI as being a less stable form of XCI, more prone to reversion. Still, a failure in balancing X-linked gene dosage deeply impairs the development and organization of early trophoblast cells[21]. This is observable either when imprinted XCI is not accomplished, as in parthenogenetic (X^MX^M) mouse embryos, or if it silences all inherited Xs, as results in androgenetic (X^PY or X^PX^P) and X^PO embryos.

Random XCI is stably maintained across the majority of cell lineages, being reverted only in specific contexts (such as gametogenesis). Of note, *Xist* does not seem to be required anymore, although its dispensability has been questioned[18,22]. Still, it continues to be transcribed and to coat the inactive X chromosome. PRC1 (responsible for H2AK119ub) and PRC2 (catalyzing H3K27me3) contribute to the maintenance of XCI across cell generations, and such role was recently shown to be necessary in the extraembryonic lineage[23]. The peri-nuclear and peri-nucleolar localizations of the inactive X chromosome have proven to be essential for the stability of the silencing state: LBR, a component of the nuclear lamina, crucially contributes to the regulatory network at the basis of XCI maintenance[20]. The peri-

nucleolar region, on the other hand, is enriched in enzymes implied in the reproduction of the inactive chromatin structure during DNA replication[6].

1.4 Interspecies variability: the study case of *Macaca fascicularis*

The mechanisms allowing XCI show fundamental interspecies variability. A recent study has suggested that XCI may not be required to exit the pluripotency state in cynomolgus monkey embryos (*M. fascicularis*)[24]. In this model, genetically closer to humans, XCI is established in random fashion for both the embryonic and the extraembryonic lineages, although with different timing. Indeed, while TE-derived cytotrophoblast cells seem to complete XCI about two days after implantation (E11), tissues originated from the ICM (amnion, epiblast, hypoblast) randomly inactivate one X chromosome later on (E15-E17). In this case, the extraembryonic lineage seems to diverge from the embryonic one by being the first to accomplish XCI (***fig. 4***).

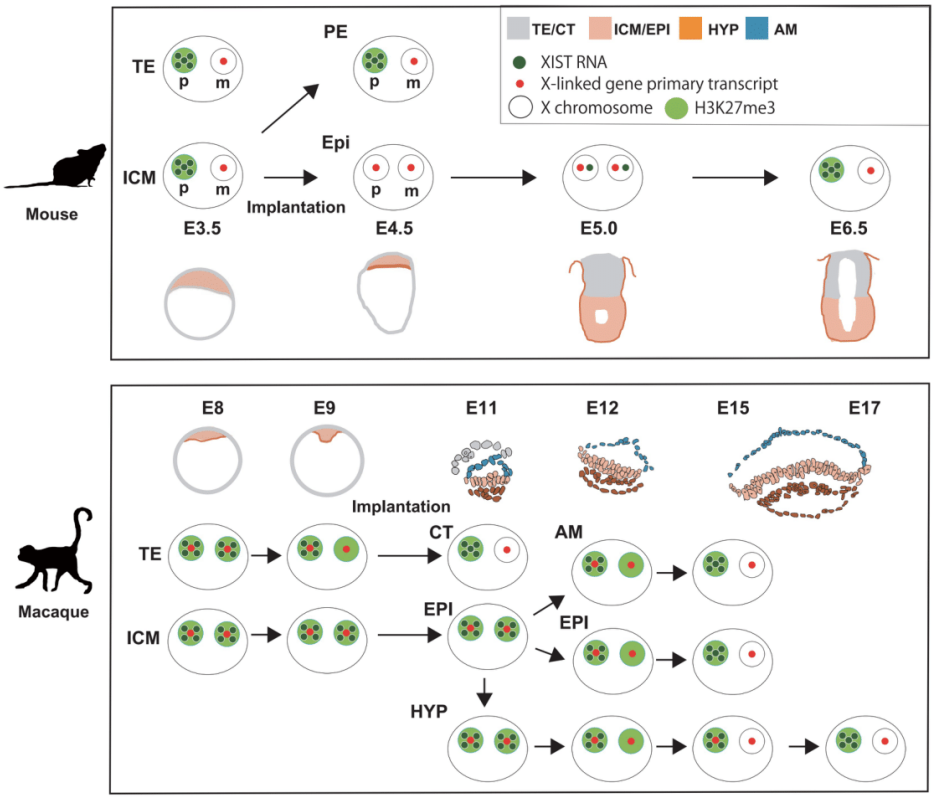

Figure 4*: Species-specific progression of XCI in different cell lineages. TE: trophectoderm; ICM: Inner Cell Mass; PE: primitive endoderm; EPI: epiblast; HYP: hypoblast; AM: amnion; p: paternally inherited X; m: maternally inherited X (Okamoto et al., 2021).*

Intriguingly, *XIST* expression resulted temporally uncoupled from both events,

being detected several days in advance (E3). Differently from the mouse, *XIST* is expressed from both X chromosomes and even in male embryos without triggering XCI. Instead, it colocalizes on both Xs together with repressive histone marks (H2AK119ub, H3K27me3). This seems to increase X chromosome compaction although X-linked transcriptional activity is not suppressed[24]. XCI starts in the trophectoderm around the implantation stage (E8-9) following the progressive downregulation of *XIST* expression and the erasure of repressive marks from the future active X. Although a lower fraction of cells from the ICM initiates XCI in the same period and with similar modalities, in this lineage the process seems to require more time to be completed. Compared to the mouse, these results suggest a different triggering modality for XCI, and the contribution of *XIST* seems to be less clear.

1.5 The state of the art for XCI in human

The study of human embryonic development is mostly based on embryos derived from *in vitro* fertilization, which can be manipulated to variable extent depending on country-specific legal terms[25]. Although being a rich source of information, it is important to consider that this kind of embryos may not be entirely reliable when studying the progression of pre-implantation human development. Indeed, they are usually conceived in a context of parental infertility, starting from potentially abnormal gametes. The frequency of aneuploidy is relatively high and their viability in culture tends to decrease with time: only 40-60% of human cleaved embryos reach the blastocyst stage *in vitro*[15].

Since 1998, it has been possible to derive and culture hESCs from human embryos[26], which share features with epiblast cells of the post-implantation blastocyst. This has opened the research field to multiple cell lines that can be virtually maintained for years. mESCs differ from hESCs in their differentiation potential, as they transcriptionally resemble the pre-implantation epiblast[27]. Mouse embryonic development seems to progress through binary and irrevocable events of fate divergence, which would explain why ICM-derived pluripotent cells fail to produce proper trophectoderm derivatives if not forced to express TE-specific transcription factors (e.g. *Cdx2*). Instead, ICM explants from late human blastocysts can still produce Trophoblast Stem Cells (TSCs) under specific culture conditions[27].

ICM-derived hESCs, defined as "primed", are in a condition poised for commitment, as evidenced by their epithelial organization, together with their transcriptional, metabolic and epigenetic profile[28]. Differently from mESCs, they correspond to post-implantation epiblast, a cellular context where XCI has already occurred. The careful manipulation of the medium composition allows to artificially reprogram primed hESCs into "naïve" hESCs[15], whose transcriptional profile resembles a more precocious stage of development. According to one recent approach[29], the calibrated use of histone deacetylase (HDAC) inhibitors, MEK inhibitor PD032591 (PD) and the human Leukemia Inhibitory Factor (hLIF) drives the initial phase of primed cell reprogramming towards a naïve profile, which is then reinforced and maintained through the addition of WNT signaling inhibitor XAV939 and aPKC inhibitor Gö6983. The addition of PD, XAV939, Gö6983 and hLIF small molecules to N2B27 basal medium defines a medium composition called "PXGL", which is currently used to for the maintenance of naïve hESCs in culture. Naïve hESCs have reverted almost, but not all, the features of XCI (***fig. 5***). For instance, the previously inactivated X tends to retain an H3K27me3 enrichment even upon reprogramming. Only one *XIST* cloud is usually observed instead of two; moreover, primed hESCs maintained in culture tend to undergo with time XCI destabilization, a culture-related phenomenon known as "XCI erosion". All this evidence hence questions the real correspondence between the pre-XCI state obtained *in vitro* and that observed *ex vivo*.

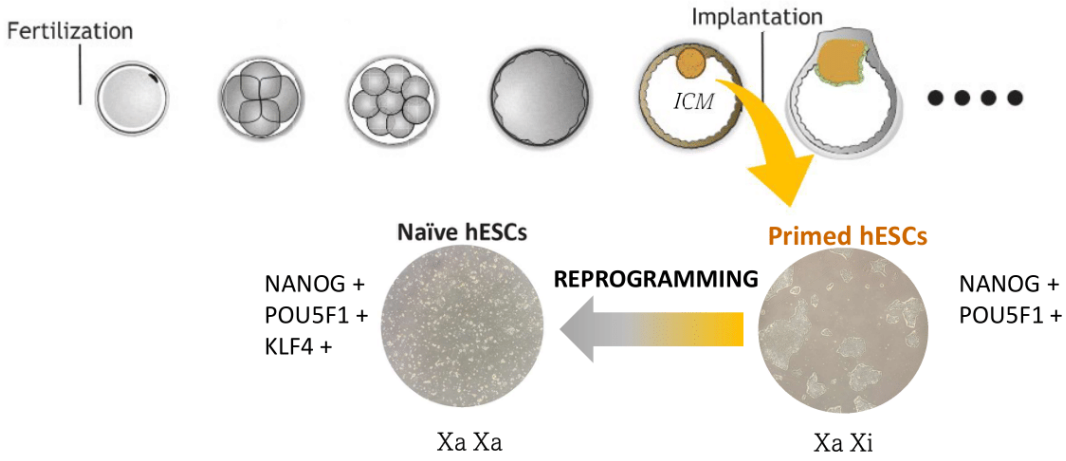

Figure 5*: Derivation of naïve hESCs through reprogrammation of hESCs in a ground ("primed") state of pluripotency. Xi: inactive X chromosome; Xa: active X chromosome (adapted from Patrat et al., 2020).*

In human pre-implantation embryos, the pre-XCI state has been defined as a condition in which both X chromosomes are active and coated with *XIST* and *XACT* lncRNAs, which start being expressed upon the activation of the zygotic genome (8-cell stage, ***fig. 6***). *XIST* is more disperse, potentially explaining why XCI is not triggered, and there is no enrichment in repressive histone marks (such as H3K27me3)[15]. Despite still being a matter of discussion, recent evidence has also corroborated a phenomenon of downregulation of X-linked gene expression preceding the actual establishment of XCI, a condition that has been termed "X chromosome dampening"[15]. *Ex vivo* studies have pinpointed the initiation of a random form of XCI at the early implantation stage (E6-7), although recent evidence suggests a slight bias towards the inactivation of the paternally inherited X chromosome in the placenta[15]. The kinetics of the process seem to show a cell lineage-dependent variability: at E8, H3K27me3 foci, indicative of XCI in somatic cells, are mostly found in TE (~ 25%) and primitive endoderm cells (~ 7.5%), while none is observed in the epiblast[15]. In this context, *XIST* coating persists only on the X chromosome to be inactivated, which also undergoes an enrichment in H3K27me3 distribution. Intriguingly, *XACT* tends to distribute on the other X, thus suggesting a potential role in shielding the chromosome from inactivation or prompting inactivation on the other X[15].

Of note, the temporal uncoupling between *XIST* expression, XCI establishment and differentiation points to a crucial divergence in *XIST* regulatory network throughout the evolution of different placental mammals. Nevertheless, *XIST* regulators and effectors are still quite elusive in human. XIC appears to lead *XIST* expression regulation, although the precise role of its loci needs to be clarified. Several non-mutually exclusive models have been proposed to explain the temporal uncoupling between *XIST* expression and XCI initiation[15]: (1) dispersion and failed recruitment of *XIST* to one of the Xs; (2) lack of downstream effectors, still to be identified; (3) XACT antagonistic activity, directly or indirectly exerted; (4) *XIST* intrinsic inability to recruit interaction partners due to mis-splicing or RNA editing.

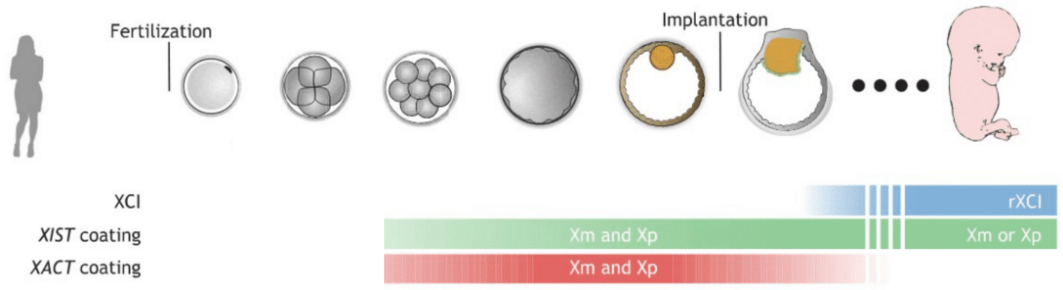

***Figure 6**: Dynamics of human XCI during early development. rXCI: random XCI; Xm: maternally inherited X; Xp: paternally inherited X (adapted from Patrat et al., 2020).*

1.6 Human Trophectoderm Cells and Trophoblast Stem Cells as a model for early extraembryonic development

With the progression of embryonic development, the trophectoderm constitutes a population of self-renewing cells, also known as Trophoblast Stem Cells (TSCs). These further specialize into different trophoblast derivatives, such as cytotrophoblast, syncytiotrophoblast and extravillous trophoblast cells, which become an integral part of the placenta[30]. These cell types play a crucial role in remodeling the organization of maternal tissues as well as the maternal immune system, thus allowing the survival of the growing embryo. After struggling with their characterization and localization, human TSCs have been recently isolated and proven to maintain self-renewal as well as their original differentiation potential in the appropriate medium conditions. On parallel, alternative methods have been proposed and adjusted to enable the obtainment of TSCs starting from naïve hESCs and reprogrammed somatic cells. Of note, the interplay of four transcription factors, being GATA2, GATA3, TFAP2A and TFAP2C, has been proposed to constitute a regulatory circuit in hESCs that would be crucial for extraembryonic commitment, as they mediate the repression of *POU5F1* while upregulating the expression of placental genes[31].

A recent comparative analysis of gene expression data from these cell lines has clarified which differentiation stage they mirror in the embryo[30]. This study has also circumscribed distinct transcriptional profiles for the extraembryonic lineage which would correspond to the progression from early to medium and late trophectoderm, flowing into the establishment of a TSC identity. This way, the expression of extraembryonic markers has been temporally limited to specific days of development: for instance, *CDX2* expression matched the medium-stage TE profile (E5-6), while *NR2F2* a more advanced period (E7-8). *GATA2* was be expressed in

a wider temporal window, comprising both TE and TSCs (E5-14). Importantly, as found in mouse, the expression of pluripotency (*POU5F1*, *NANOG*, *SOX2*) and trophectoderm markers (*CDX2*, *NR2F2*, *GATA2*) resulted mutually exclusive.

A publication from *Guo et al.* has proposed a novel approach to derive TE cells from naïve hESCs by playing with specific signaling pathways[27]. The protocol is based on a culture medium formulation named "PDA83", consisting of N2B27 basal medium supplemented with PD0325901 (MEK inhibitor) and A83-01 (NODAL/TGF-β signaling inhibitor). Culturing naïve hESCs in this medium for 5 days strongly deranged cell behavior, promoting the acquisition of a TE-specific transcriptional signature. In terms of transcription, most day-5 cells showed proximity to the profile of trophectoderm (pre-CTB) and cytotrophoblast (CTB) embryo-derived tissues (***fig. 7***). Importantly, such TE cell cultures could be employed for the generation of expandable TSCs.

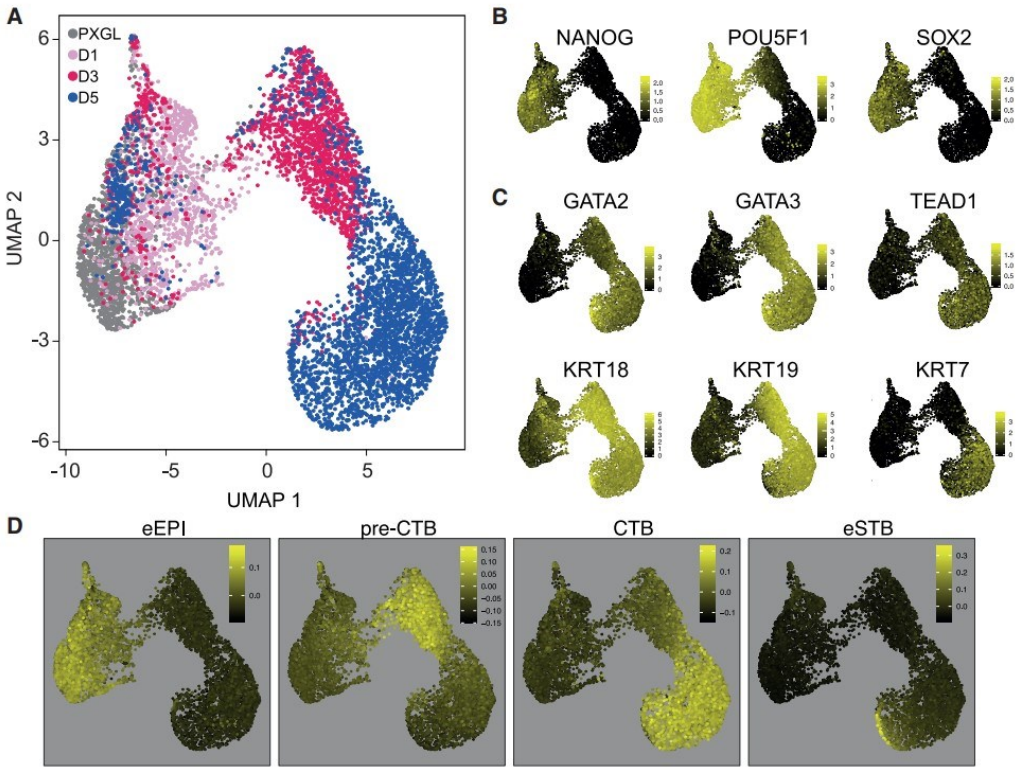

Figure 7*: Single cell analysis from Guo et al. publication (2021). A: UMAP representation of single cell RNA-sequencing data collected from naïve hESCs (PXGL), day 1 (D1), day 3 (D3) and day-5 (D5) cells cultured in PDA83 medium. B: normalized expression of pluripotency markers in (A). C: normalized expression of selected TE and trophoblast markers in (A). D: normalized expression in (A) of genes enriched in the indicated human embryonic stages (data collected from extended cultures of human embryos). eEPI: E6-E8 epiblast; preCTB: E6-E7 trophectoderm; CTB: cytotrophoblast; eSTB: early syncytiotrophoblast.*

1.7 Presentation of the experimental plan: aims and preliminary results

Recent evidence has shown that XCI is completed during a 30 days-long specification protocol of naïve hESCs into TSCs (*host laboratory unpublished data*). Importantly, shutdown of *XIST* expression appears to impair the process *in vitro* (***fig. 8***), implying the pivotal role of this lncRNA in the establishment of XCI. What is more, XCI disruption has been associated to an increased failure in the derivation of TSCs, with diffuse cell death over time (*host laboratory unpublished data*). This indicates XCI as a necessary event for the proper progression of the extraembryonic fate.

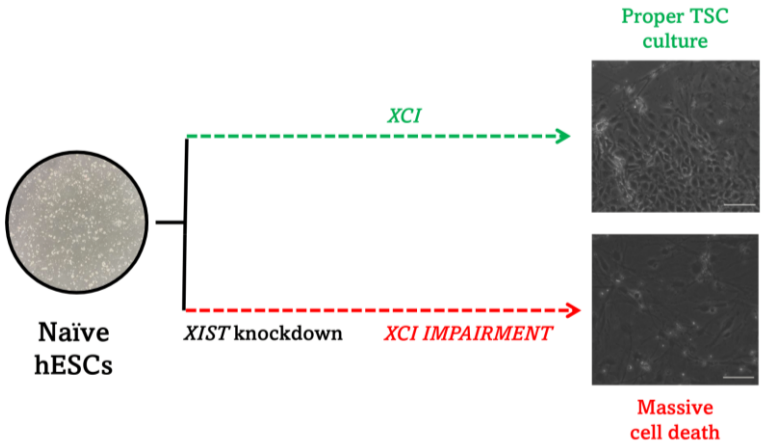

Figure 8*: XIST repression affects XCI establishment and cell survival throughout a 30-day long TSC conversion protocol (host laboratory unpublished data).*

Guo et al. have opened the possibility to explore the early events of X-linked gene expression regulation during the extraembryonic fate commitment. The present project digs into the relevance of XCI in hESCs during the exit from a naïve pluripotency state towards the acquisition of a trophectoderm-like profile. We aimed to clarify (1) whether XCI occurs in this time window, preceding the establishment of TSCs, and (2) the effect of *XIST* repression on the phenotype, in terms of both X chromosome dynamics and efficiency of TE specification.

To this aim, we used H9 (WA09) hESC cell line, originally expanded from the ICM of a female human blastocyst where XCI had already occurred[26]. Three primed H9 subclones (namely 1, 2, 3) had been successfully engineered by randomly introducing in the genome a doxycycline-inducible CRISPRi repression system organized in two transgenic cassettes (***fig. 9***). The insertion was driven by piggyBac

transposase, expressed from a third unrelated construct. In these subclones, upon pharmacological treatment with doxycycline (DOX), a catalytically-inactive Cas9 fused with a KRAB repressive domain is expressed and guided towards the promotor of *XIST* locus by a specific gRNA, allowing for local transcriptional silencing[32]. Primed hESC CRISPRi subclones had been reprogrammed towards a naïve state of pluripotency by adapting current protocols[29].

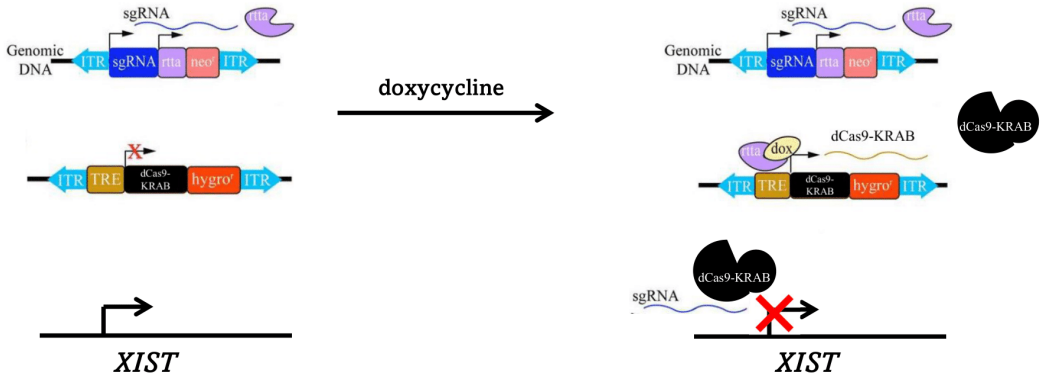

Figure 9: *Schematic representation of the CRISPRi strategy, based on the insertion in the genome of two constructs. The sgRNA and the reverse-tetracycline transactivator (rtTA) are constitutively expressed. In presence of doxycycline, rtTA triggers the expression of dCas9-KRAB fusion protein, which will recruit the sgRNA and repress XIST expression via local deposition of H3K9me3, a histone post-translational modification triggering transcriptional silencing (adapted from Schertzer et al., 2019)*

2. MATERIAL AND METHODS

2.1 Cell culture

H9 naïve hESCs were cultured on a layer of mitotically inactivated Mouse Embryonic Fibroblasts (iMEFs) in Thermo Scientific™ Nunc™ cell culture plates. They were maintained in PXGL medium at T = 37°C under hypoxic conditions (5% O_2, 5% CO_2) and subjected to daily medium changes. Every 3-5 days they were passaged at 1:1 to 1:4 dilution rate following dissociation to single cells with Accutase™ or TrypLE™ treatment for 5-10 minutes. ROCK inhibitor Y-27632 (Merck Chemicals; final concentration: 1 μg/mL) was added after every split to improve the cell survival rate.

iMEFs where thawed few days to few hours before naïve cell split, counted and plated on 0.1% gelatin-coated wells at a cell density spanning from ~ 90K (24-well plate) to ~ 350K cells per well (6-well plate). MEFs were incubated at T = 37°C in normoxic conditions (20% O_2, 5% CO_2) until the naïve hESC passage.

The pharmacological treatment was initiated at least 10 days before launching a protocol of TE specification. Doxycycline was directly added to the medium at a final concentration of 1 μg/ml and administered everyday throughout the experimental workflow. Its effect on *XIST* expression was validated and tested via RNA-FISH.

Culture media composition

- MEF medium

Component	Final concentration	Source and catalog ID
DMEM	100%	Life Technologies, 11965092
FBS	10%	Dutscher, P04-96650
Penicillin, Streptomycin	1X (100X stock)	Life Technologies, 15140122

- N2B27 medium (to be filtered and stored at 4°C for approximately one month)

Component	Final concentration	Source and catalog ID
DMEM/F-12 medium	50%	Sigma, D8437-500ML
Neurobasal medium	50%	ThermoFisher, 21103049
L-glutamine	2 mM	ThermoFisher, 25030024
2-mercaptoethanol	100 μM	Life Technologies, 31350010
N2 supplement	1X (200X stock)	ThermoFisher, 17502048
B27 supplement	1X (100X stock)	ThermoFisher, 17504044

- PXGL medium (to be filtered and stored at 4°C for approximately one week)

Component	Final concentration	Source and catalog ID
N2B27 medium	100%	*Made in-house*
PD0325901	1 μM	abcr, AB 253775
XAV939	2 μM	Cell Guidance Systems, SM38-200
Gö6983	2 μM	Bio-Techne, 2285/10
human LIF	10 ng/mL	PeproTech, AF-300-05

- PDA83 medium

Component	Final concentration	Source and catalog ID
N2B27 medium	1:1	*Made in-house*
PD0325901	1 μM	abcr, AB 253775
A83-01	1 μM	PeproTech, 9094360

Media were warmed to room temperature (15-25°C) before use.

The specification towards a trophectoderm-like identity was induced according to the following protocol, adapted from the publication of *Guo et al.*[28]

- *Day -1*: naïve hESC are passaged and distributed onto Matrigel-coated wells. For 24 well plates, different cell densities were assessed: 50,000 and 100,000 cells per well, approximately equivalent to 25,000 and 50,000 cells/cm^2 respectively. For 6 well plates, the starting cell density was set at 200,000 cells per well (~ 25,000 cells/cm^2).
- *Day 0*: the medium is switched from PXGL (/PXGL + DOX) to PDA83 (/PDA83 + DOX).
- *Day 1-5*: PDA83 medium is changed daily. During this time span, cells tend to undergo visible morphological changes and organize epithelial-like structures.

Importantly, cells under TE specification were incubated in hypoxic conditions throughout the experimental process. After day 5, TE-like cells were no longer maintained in culture.

For RNA-FISH and immunofluorescence analysis, hESCs cells were grown on Matrigel-coated coverslips in 24 well plates. For FACS, in order to obtain a sufficient number of cells, 6 well plates were used. For the analyses, when possible, three CRISPRi subclones were assessed in parallel together with a wildtype control, under both treated (DOX+) and untreated (DOX-) conditions. The karyotypic profile of each subclone had been validated in the primed state.

Bright-field images were taken using Olympus CKX41 light microscope (4X) and Leica DM IL LED inverted epifluorescence microscope (10X).

2.2 RNA-FISH probe synthesis by nick translation

The vectors carrying the DNA probes for RNA-FISH were extracted from stock cultures of bacteria stored at -70°C. After qualitative and quantitative assessments of the minipreps, each vector (1 µg) was subjected to nick translation provided with a mix of canonic dNTPs (dATP, dGTP, dCTP 0.2 mM each, dTTP 0.1 mM) and fluorescent dUTPs (1 mM). The reaction was incubated at 16°C overnight and the fluorescent probes stored at -20°C.

Vector	Target gene	Species	UCSC insert name/position	Resistance marker
Plasmid pUC18	*XIST*	*Homo sapiens*	chrX:73,039,838-73,049,837	Ampicillin
BAC pBACe3.6	*POLA1*	*Homo sapiens*	RP11-1104L9	Chloramphenicol

2.3 RNA-Fluorescence *In Situ* Hybridization (RNA-FISH)

RNA-FISH allows to detect the presence and localization of RNAs of interest. In this context, the technique was crucial to determine the monoallelic rather than biallelic expression of X-linked genes.

Cells were fixed at room temperature in 3% PFA/1X PBS for ten minutes, washed at least once in PBS 1X and permeabilized on ice for seven minutes in freshly made CSK/0.5% Triton X-100 supplemented with VRC RNase inhibitor (1:100). The samples were dehydrated with ethanol washes at increasing concentration (70%, 90% and 100%) and finally incubated at T = 37°C overnight with the hybridization mix containing the probe(s) of interest. The following day cells were washed three times in 50% Formamide/2X SSC at pH=7.2 and three times in 2X SSC to remove the Formamide. Each coverslip was mounted on glass slides using Vectashield® with DAPI to stain the nuclei.

2.4 Immunofluorescence

Cells were fixed at room temperature in 3% PFA/1X PBS for ten minutes, washed at least once in PBS 1X and permeabilized on ice for seven minutes in freshly made 1X PBS/0.5% Triton X-100. After at least two washes in PBS 1X, non-specific binding of the antibodies was prevented with 15 minutes incubation in 1X PBS/1% BSA blocking buffer. Coverslips were incubated at room temperature with the solution of primary antibodies in a humid chamber for 45 minutes, washed at least once in PBS 1X and incubated with the solution of fluorescent secondary antibody in a dark, humid environment for another 40 minutes at room temperature. After at least one final wash in PBS 1X, the coverslips were mounted in Vectashield® with DAPI on glass slides. The following primary antibodies and secondary antibodies were used for the analyses. Of note, secondary antibodies were erroneously used at a concentration ten times higher than recommended.

Primary antibody	**Dilution**	**Source and catalog number**	**Secondary antibody**	**Dilution used**	**Source and catalog number**	**Fluorophore**
Anti-NANOG	1:1000	Abcam 21624	Chicken anti-rabbit IgG	1:100 (1:1000 dilution recommended)	Invitrogen A21443	Cy5
Anti-GATA3	1:100	Invitrogen 14-9966-82	Donkey anti-rat IgG	1:100 (1:1000 dilution recommended)	Invitrogen A21208	Alexa Fluor 488

Confocal images were acquired at the Leica DMI6000B confocal microscope through MetaMorph software, setting 20X-40X oil immersion objectives with a step of 0.3-0.4µm on the Z axis. Fiji software was used for image processing.

2.5 RT-qPCR

Cells were lysed with TRI Reagent™ (0.5-1 ml) and total RNA was isolated according to the Sigma-Aldrich technical bulletin. 500 ng of template RNA per condition were subjected to reverse transcription using random hexamers (50 μM) and Superscript® IV Reverse Transcriptase. The enzyme was not added in the negative controls (RT-). The incubation timings were adapted from Invitrogen User Guide, increasing the reaction time at 50-55°C up to 30 minutes.

SYBR green PCR master mix was used for the qPCR experiment. For each condition tested (subclone X; gene N), two technical replicates were provided. The average Ct was considered for further analysis, calculating 2^{-Ct} and normalizing it to the corresponding value obtained for β-actin. *GAPDH* and *VIMENTIN* were tested as well, the latter being an extraembryonic mesoderm marker gene, not expressed in TSCs, naïve and TE-like cells. RT- and H_2O were used as negative controls.

2.6 Fluorescence-Activated Cells Sorting (FACS)

Two trials were made to sort (in the first case) and analyze (in both cases) TE-like cells based on *ENPEP* expression. Cells were dissociated with TrypLE™, collected and washed with PBS 1X/0.5% BSA (FACS buffer). One million cells were then incubated with ENPEP fluorescent antibody (BD Biosciences, BD Optibuild™ BV421 Mouse Anti-Human CD249) at 4°C for 30 minutes in a dark chamber. Afterwards, cells were washed with FACS buffer one last time and analyzed. Naïve hESCs and hTSCs were used as reference.

2.7 Statistical analysis

Regarding RNA-FISH, IF and ENPEP-based cell sorting results, the difference found in the proportion of *XIST*, *POLA1*, *GATA3*, *NANOG* and *ENPEP* expression categories was statistically assessed on Prism using a χ^2 test (significance for $p < 0.05$). For each group (DOX-, DOX+), the countings from all subclones were pooled together. Importantly, this approach allows to test whether a difference between two conditions is significant without providing information on how single

subclones contributed to said difference. To gain more information, a χ^2 test was also used to test the homogeneity in the distribution of values within each group (DOX-, DOX+).

As for the RT-qPCR data, the difference in terms of RNA levels was evaluated using one-tailed (for *XIST*) or two-tailed (for all the other genes) unpaired t-tests on the Prism platform. In order to choose the most appropriate test, the Shapiro-Wilk test was first employed on R to check for the intra-group normal distribution of the data. If normality could not be supported ($p < 0.05$), the Mann-Whitney non-parametric test had to be used. When normally distributed, the variance equality between groups was further assessed on R to choose whether to use a parametric t-test (which assumes the same standard deviation among the two tested groups) or the Welch's t-test variant (which is not limited by the variance equality assumption).

3. RESULTS

3.1 Optimization of naïve hESC conversion to trophectoderm-like identity

We first set out to optimize the conditions for conversion of naïve hESC to trophectoderm (***fig. 10A***) by using a different starting concentration of plated naïve cells. All three subclones harboring the transgenes necessary for *XIST* repression were tested at once. The initial trials of TE conversion produced different outcomes depending on the initial cell density set at day 0. When naïve hESCs were seeded on Matrigel-coated coverslips with a dilution of approximately 25,000 cells/cm^2, the phenotype progressed similarly to what previously described by *Guo et al*. Human ESCs shifted from a roundish to a more flattened morphology, from being organized into small and isolated colonies, typical of the naïve state, to the constitution of epithelial-like larger colonies (***fig. 10B***, upper row). On the other hand, if plated at twice the concentration (~ 50,000 cells/cm^2), all subclones tended to quickly reach high confluency and undergo diffuse cell death. Morphologically speaking, the phenotype differed from what expected (***fig. 10B***, bottom row): no evident epithelial-like organization was observed even though naïve cell colonies tended to flatten and enlarge, which is a typical sign of differentiation, as well as merge and overgrow on top of each other when reaching high peaks of cell density. These preliminary results questioned the actual efficiency of the process and contributed to the decision of maintaining an initial cell concentration of ~25,000 cells/cm^2 for subsequent TE conversions.

Nonetheless, both setups provided an upregulation of TE-specific marker genes, which was tested via immunostaining and RT-qPCR. The expression of GATA3 and NANOG transcription factors was assessed by immunofluorescence to check the proportion of TE-like and naïve-like cells respectively (***fig. 10C***). Despite technical issues with the anti-NANOG antibody, it was possible to highlight (1) a general increase in *GATA3* expression for *day-5* TE cells compared to the wildtype naïve condition (***fig. 10C***, right panel) and (2) the mutual exclusivity between NANOG and GATA3 staining. By RT-qPCR (***fig. 10D***), a clear trend emerged among all CRISPRi subclones, which consisted of the downregulation of pluripotency and naïve-specific markers (*NANOG*, *POU5F1*, *KLF4*), with the concomitant upregulation of *GATA2, GATA3* and, even if more moderate, *NR2F2*. Of note, *NR2F2* had not been previously assessed in the reference paper (*Guo et al*.).

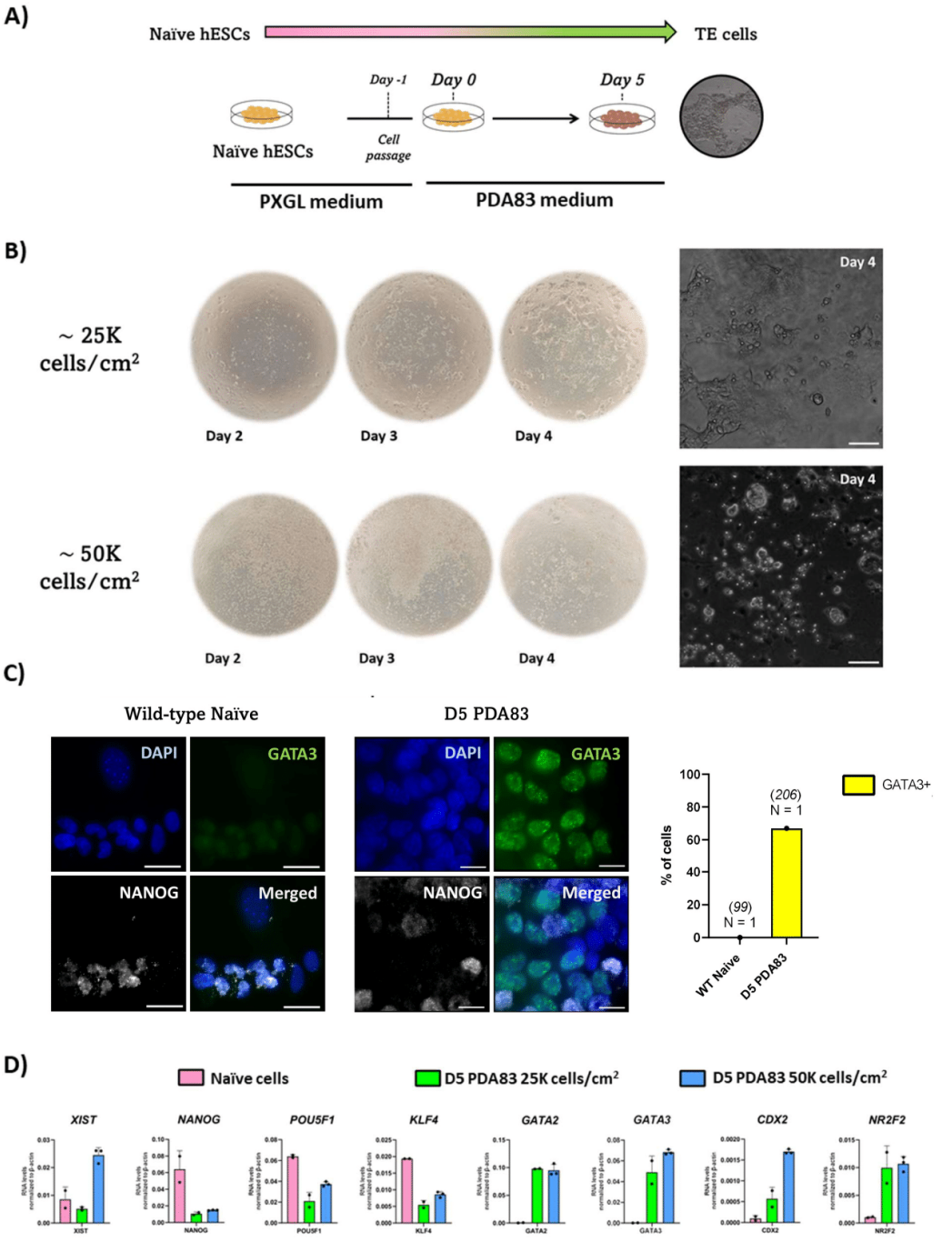

Figure 10: Optimization of naïve hESC conversion to trophectoderm-like identity.

*A) Scheme of the TE conversion from naïve hESCs. Images of TE cells are from Guo et al. publication[27]. **B)** H9 cells under TE specification launched at different starting cell density and documented from day 2 to day 4 (4X magnification). A closeup of the cell state at day 4 is here reported as representative for both the wildtype and the CRISPRi subclones. Scale bar = 100 µm. **C)** H9 naïve hESCs versus day-5 of TE specification, both immunostained to reveal GATA3 and NANOG expression. Scale bar = 20 µm. **D)** RNA expression levels normalized to β-actin for a panel of marker genes used to compare, in this context, naïve and day-5 TE cells obtained at two different starting cell densities (~25K cells/cm², ~50K cells/cm²). TE-50K data were obtained from an independent experiment.*

Although higher in TE cells compared to naïve counterparts, relatively low expression was detected for *CDX2* (medium-stage TE). The result for *CDX2* is coherent with previous findings, which showed a transitory peak of expression within days 2-3 of TE specification[27]. Altogether, RNA expression data revealed a similar transcriptional signature shared by CRISPRi subclones independently of whether the TE specification protocol started from 25,000 rather than 50,000 cells/cm^2: in both conditions, cells showed a change in morphology supporting (1) exit from pluripotency and (2) commitment to the extraembryonic fate. As a side note, *XIST* expression did not seem to be equally homogeneous, questioning the actual similarity of the normalized gene expression levels between the two protocols (25K cells/cm^2, 50K cells/cm^2). The observed difference in *XIST* expression may be related to technical issues occurred during the setup of the RT-qPCR.

2.1 Induction of *XIST* expression via doxycycline treatment is functional in trophectoderm cells

Two culture conditions were set for each CRISPRi subclone: the "DOX-", in absence of drug treatment, and the "DOX+", supplemented with doxycycline (***fig. 11A***). The pharmacological treatment was maintained for at least 10 days to assure an efficient *XIST* knockdown in naïve hESCs. After this period, both the "treated" and the "untreated" subclonal lines were subjected to TE specification in parallel. We adapted the 6-days long protocol from *Guo et al.* publication, consisting in the setup of a feeder-free naïve culture (*day -1*) followed by a switch of medium from PXGL to PDA83 on the next day (*day 0*). The conversion was stopped at day 4 or day 5, proceeding with downstream evaluations. *XIST* knockdown has been verified in both naïve and *day-5* DOX+ conditions via RNA-FISH (***fig. 11B***).

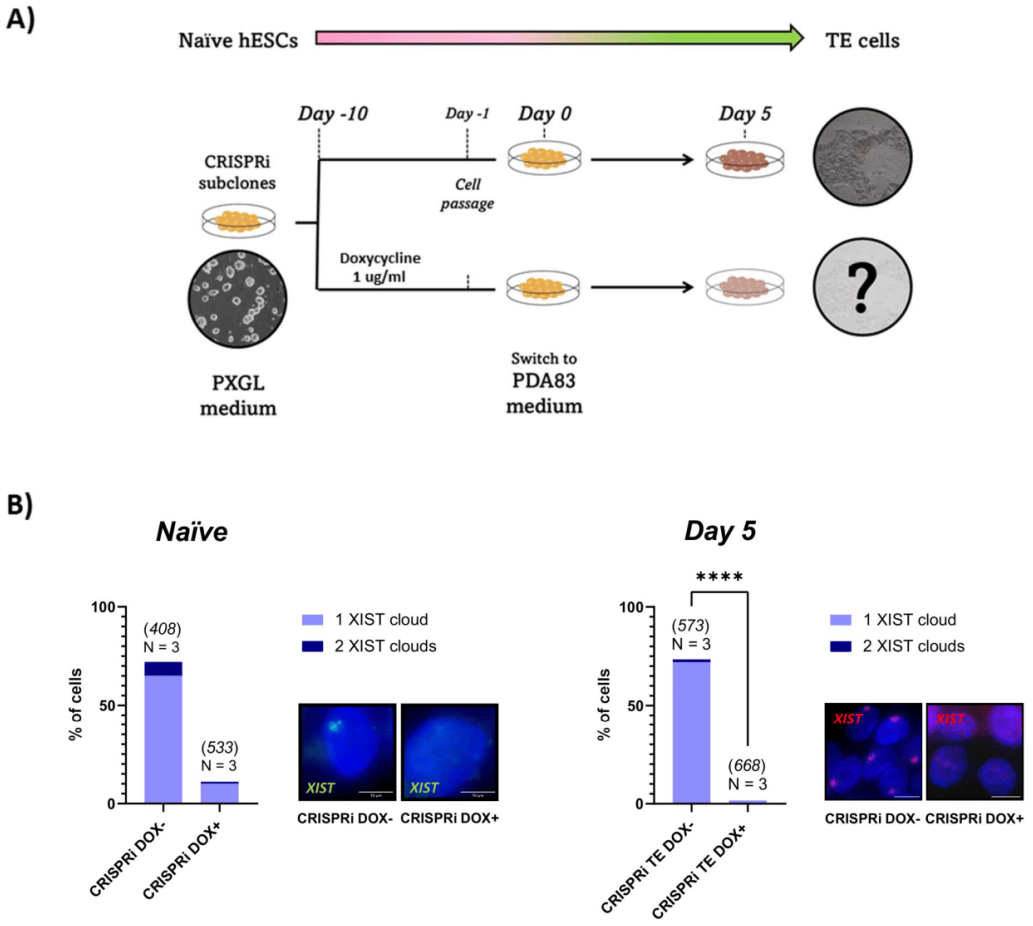

Figure 11: XIST repression is maintained during TE conversion.

*A) Scheme of the experimental strategy. TE specification was launched for all three CRISPRi subclones in parallel. Images of naïve (left) and TE cells (right) are from Guo et al. publication[28]. **B)** XIST expression and localization defined as one, two or no clouds per cell. Both naïve (left) and day-5 TE (right) CRISPRi clones were assessed (N = 3). At least 100 cells were counted per condition. A χ^2 test was performed on pooled counting (statistical significance for $p < 0.05$, **** notation when $p < 0.0001$). Scale bar = 10 µm.*

2.2 Characterization of the *XIST* knockdown phenotype on TE-like converted cells

The dynamics of XCI were checked by RNA-FISH, aiming to assess the expression state of *POLA1* that is known to switch from biallelic (pre-XCI) to monoallelic expression after XCI (***fig. 12A***). The analysis showed a significant increase in the frequency of *POLA1* monoallelic expression for the untreated CRISPRi group compared to DOX+ counterpart. These results suggest that (1) XCI is indeed initiated in the trophectoderm and (2) it is deeply affected by *XIST* expression, as *POLA1* biallelic expression was predominantly maintained upon *XIST* knockdown, presumably by a failure of XCI to proceed in absence of *XIST*.

The next step was to assess whether *XIST* knockdown, and likely XCI impairment, could affect the capacity of naïve hESCs to exit the pluripotency state and acquire a trophectoderm identity. Unexpectedly, the analysis of *GATA3* expression in converted cells via IF gave contrasting results (***fig. 12B***). Although all subclones seemed to exit the pluripotency state by losing *NANOG* expression, a relevant inter-clone variability resulted for both conditions when considering *GATA3* (DOX-, DOX+). This time the differences were even more pronounced than what previously observed among data of IF, RNA-FISH and RT-qPCR. Intriguingly, while for two subclones the relative efficiency of TE specification was comparable among the treated and untreated conditions (subclones 1 and 3), one stood out showing a larger fraction of GATA3 positive cells in the absence of doxycycline. The statistical analysis performed on pooled cell counting (χ^2 test, as suggested by the in-house biostatistician) highlighted a strongly significant difference between DOX- and DOX+ outcomes. Nonetheless, this difference was contributed only by subclone 2.

In the results described so far, we have shown that all subclones of naïve hESCs are able to undergo *GATA3* upregulation although with different efficacy. We have also shown a tendency of *XIST* expressing cells to initiate XCI along the process with higher frequency compared to the *XIST* knockdown condition. Although the results suggested XCI to be a dispensable event for TE fate determination, we wanted to explore this possibility with higher detail. To check for the presence of any correlation between the frequency of XCI occurrence and the acquisition of a TE-like identity, we combined Fluorescence-Activated Cell Sorting (FACS) with RNA-FISH. CRISPRi subclones 2 and 3 (each in the treated and untreated condition) were subjected to TE specification and collected at day 5.

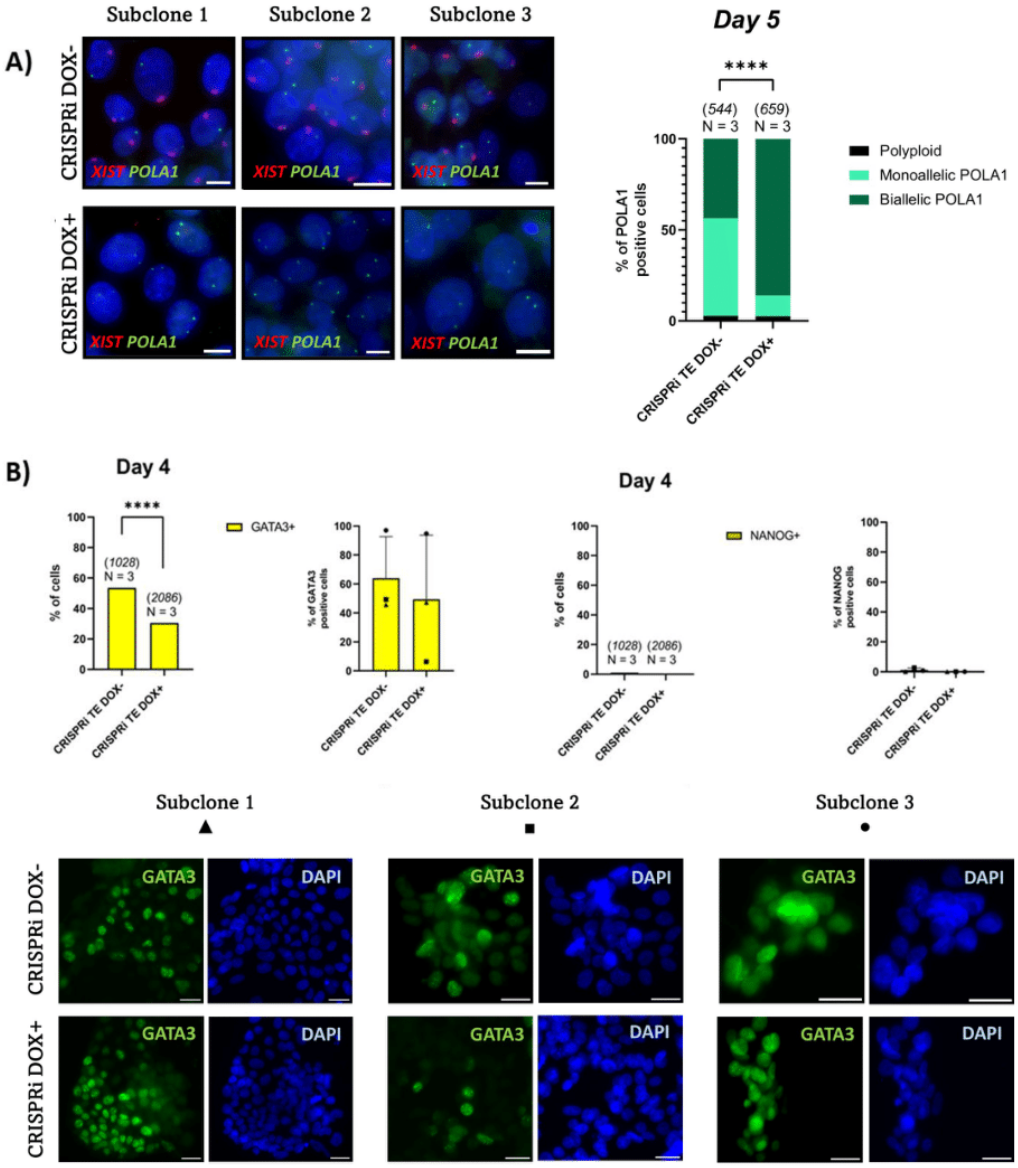

Figure 12: XIST repression impairs X-chromosome inactivation and leads to heterogenous TE conversion.

A) *On the left, representative images from an RNA-FISH experiment performed on CRISPRi clones (DOX+, DOX-). Scale bar = 10 µm. On the right, the distribution for the percentages of monoallelic, biallelic and multi-allelic POLA1 extrapolated from pooled counts. More than 100 cells were counted per condition. A χ^2 test was employed when comparing DOX- and DOX+ groups (significance for $p < 0.05$).* ***B)*** *percentage of GATA3 and NANOG positive cells among day 4 TE CRISPRi subclones, with more detail provided on single subclone behavior in the adjacent histograms. Below, representative GATA3 immunostaining pictures for all conditions tested. Scale bar = 20 µm. More than 100 cells were counted per condition. A χ^2 test was used for GATA3 cumulated counts, significance set for $p < 0.05$.*

Throughout this conversion procedure, a difference in cell density became evident between DOX+ and DOX- subclones despite having seeded the same initial cell concentration, with the untreated conditions showing a lower confluence compared to their counterparts (***fig. 13A***). Intriguingly, a similar difference between the DOX- and DOX+ conditions had been noticed in previous TE specification launches, yet not really explored and sometimes attributed to technical mistakes occurred when passaging and counting naïve cells. Morphologically speaking, while subclone 3 tended to organize in large epithelial-like colonies in both DOX+ and DOX- conditions, subclone 2 under doxycycline showed a less defined colony organization with smaller cells. These observations were recapitulated in a second independent TE conversion launch.

For FACS analysis, cells were sorted in ENPEP+ and ENPEP- cells. ENPEP glutamyl aminopeptidase, also known as APA, is an integral membrane protein shown to be upregulated in trophoblast progenitors[33]. Of note, it had already been successfully exploited to analyze and sort TSC subpopulations in culture. In the control condition, here given by TSCs, the ENPEP+ and ENPEP- populations are represented as two clearly distinguishable yet partially overlapping curves, with ENPEP+ cells being a minor yet not neglectable fraction (35,4%, ***fig. 13B***). Unexpectedly, despite two FACS experiments, the majority of TE-like cells resulted ENPEP negative, with a percentage of positivity never superior to 8%. Untreated subclones seemed more prone to undergo ENPEP upregulation, as the percentages of ENPEP-positive cells ranged from 1.0 to 7.3% against 0.0-0.9% positivity found in treated subclones (***fig. 13C***). ENPEP+ and ENPEP- cells were collected for each condition, fixed and permeabilized directly on glass slides, ready to be evaluated by RNA-FISH. Unfortunately, too few cells were detected in each case, denying any possibility of counting. This result was most probably due to a combination of relatively few starting cells and technical issues.

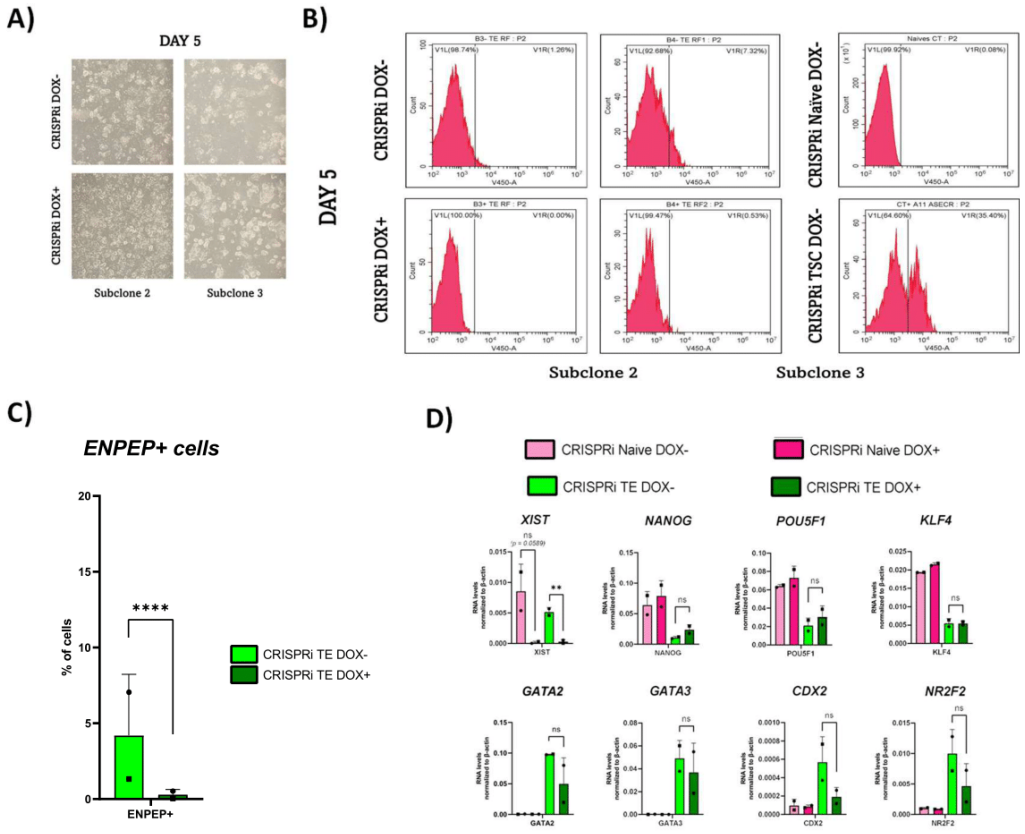

Figure 13: XIST repression perturbs derivation of TE-like cells.

*A) Brightfield images of day-5 converted CRISPRi subclones cultures. B) FACS performed on the same cells as in (A). V1L panel groups ENPEP- single cells, while ENPEP+ single cells are shown in V1R panel. C) Graphs show the distribution of ENPEP+ and ENPEP- cells in a subset of 20,000 cells (**** notation for $p < 0.0001$). D) RT-qPCR, β-actin used as normalizer gene. Statistical analysis was performed with a t-test, either non-parametric or parametric depending on the distribution of the data (** notation for $p < 0.01$; ns for $p \geq 0.05$).*

Considering the general high cell confluency observed at day 5, we first hypothesized that the antibody had been used at a too low concentration causing the signal to be diluted out. For this reason, the second FACS was performed starting with a lower number of cells (≤ 1,000,000 per condition) so to guarantee a better staining. However, even in this second case the results recapitulated what first observed, meaning an almost complete absence of ENPEP+ cells. This result questions the actual identity of the cells that were tested and, potentially, the impact of *XIST* suppression in reaching it. A technical issue may underlie this result. One alternative explanation could be that ENPEP upregulation occurs at a step not fairly represented by *day-5* H9 GATA3+ cells, at least concerning CRISPRi subclones. Either way, ENPEP upregulation was shown to be significantly impaired in the absence of *XIST* expression.

The expression of pluripotency and trophectoderm markers was assessed by RT-qPCR on subclone 2 and 3 (DOX+, DOX-) in both the *naïve* and *day-5 TE* states (***fig. 13D***). Of note, none of the trophectoderm and trophoblast expression markers significantly differed (despite a tendency) between DOX+ and DOX- conditions, implying the dispensable role of *XIST* expression for their upregulation throughout the differentiation process. Interestingly, the subclones showed again recurrent variability in the expression of TE markers, especially considering *GATA3*, *CDX2* and *NR2F2*. This intra-group variability, together with the low number of samples, were probably the main reasons behind the non-significant difference found between DOX- and DOX+ groups, including that for *XIST* expression in naïve samples. That said, the impact of *XIST* knockdown is hard to infer and needs careful consideration given the fact that the tested CRISPRi subclones, ideally biological independent replicates, were not always behaving homogeneously.

4. DISCUSSION

The extraembryonic lineage is the first to be specified through the constitution of the trophectoderm tissue at the blastocyst stage. Although being conserved in marsupials (Metatherians)[34], it soon acquires distinctive features in placental mammals (Eutherians) as it has evolved to mediate implantation in the uterus and the organization of the placenta together with maternal tissues, allowing intra-uterine gestation. Intriguingly, *XIST*-directed XCI was found to be unique to Eutherians as well and required for proper embryonic development.

From the body of data so far collected, we propose XCI to initiate in medium-late stage TE-like cells and to be a dispensable event for TE commitment, even if it would fundamentally contribute to cell survival and fate determination at later stages (***fig. 14***). In this context, the expression of *XIST* proved to be necessary for XCI initiation.

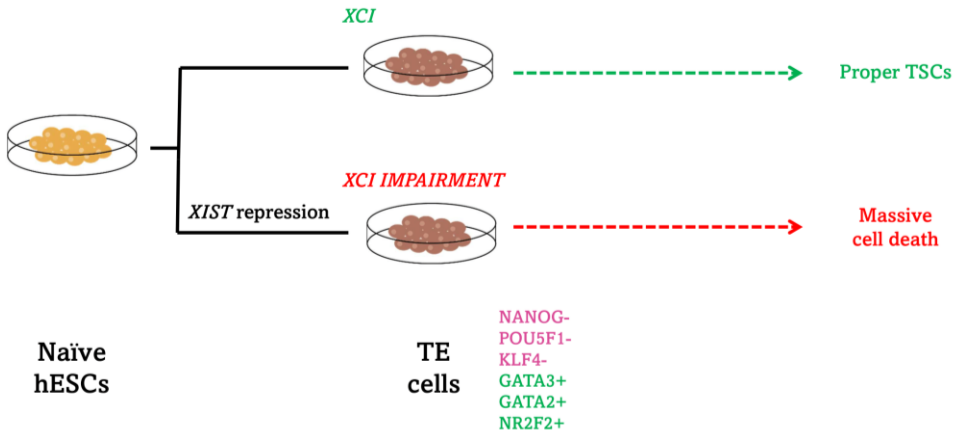

***Figure 14**: Model for XCI role in the early extraembryonic development.*

It is true that trophectoderm fate acquisition and XCI establishment were assessed in two independent experiments (IF and RNA-FISH, respectively), thus an ambiguity remains on whether and to what extent XCI is preferentially initiated in TE-like cells compared to non-TE-like cells. Unfortunately, the combination of FACS and RNA-FISH was unsuccessful, yet there is room for technical improvements. Available protocols also allow to sequentially combine IF with RNA-FISH, thus providing a parallel way to corroborate the results. Finally, bulk RNA sequencing emerged as a potentially interesting tool for the study of XCI progression. Indeed, among gene expression data (useful to determine, for instance, the extent of trophectoderm commitment), it would be possible to extract the

relative expression profile of X-linked genes. Analysis on primed H9 hESCs have shown that XCI is skewed towards one X, as during clonal expansion this cell line conserves epigenetic memory of the inactive X originally present in the few ICM cells from which it was generated. Recent data have suggested a skewed re-inactivation of the same X during the derivation of TSCs from naïve hESCs (*host laboratory unpublished data*). Given the presence of allele-specific SNPs on X-linked loci, it would be theoretically possible to measure the relative RNA expression levels from both X chromosomes and relate the result to the progression of X chromosome inactivation with bulk RNA-sequencing. Importantly, this technique provides an average level of RNA expression extrapolated from the entire cell population. Single-cell RNA sequencing could be integrated to shed light on the transcriptional homogeneity and individual cell identity within the sample.

H9 hESC line was able to exit pluripotency and acquire trophectoderm features (most notably *GATA2* and *GATA3* upregulation) independently of *XIST* expression and XCI. To reach this conclusion, we used three different independent subclones able to suppress *XIST* expression under doxycycline treatment. Intriguingly, an evident inter-clone variability emerged throughout the experiments, with particular reference to IF and RT-qPCR assays, questioning the real impact of *XIST* downregulation. Importantly, no clear link was identified between the variability of *GATA3* expression and that of *POLA1* monoallelism for each subclone. Once excluded technical biases, it would be useful to further deepen the biological basis of such variations, which may be symptomatic of players in the pluripotency state maintenance or XCI dynamics being differentially expressed. Considering that the same culture and experimental conditions were set, a plausible cause of such heterogeneous profiles could be side effects dependent on the insertion site of the CRISPRi constructs used to build the *XIST* CRISPRi inducible system. Given that piggyBac transposase tends to target euchromatic regions, the insertion may have altered the regulative or the coding sequence of gene(s) involved in the specification process.

Considering the RT-qPCR results, the DOX+ condition showed a tendential lower efficiency in the upregulation of TE markers, which was not statistically supported also due to the limited number of samples and the presence of inter-clone variability. For sure, replicates of this experiment are needed to provide enough evidence for a more robust statistical analysis. Even if slight, a statistically supported difference

in the efficiency of trophectoderm commitment could be determinant for proper embryo development and survival, especially at this early pre-implantation stage.

All the results provided, we suggest that *XIST* downregulation may impact aspects of TE specification such as cell viability and/or cell proliferation rate, as well as ENPEP expression. Surprisingly, doxycycline-treated subclones displayed a tendency to reach a higher cell density compared to the untreated counterparts, opposed to what observed during the TSC specification protocol. This may imply a role of XCI in the regulation of the proliferation rate, for example, to guarantee an equilibrium that is compatible with cell survival in more advanced stages of development. First, these observations would need to be quantified through proliferation and cell survival assays as well as statistically revised. If confirmed, they would pave the way to the search for and functional validation of X-linked genes involved in the phenotype, with *XIST* expression being an interesting target for the manipulation of XCI establishment. At day 5 of TE conversion, more than 50% of cells had undergone *POLA1* monoallelic inactivation. In light of these results, it is conceivable that this timeframe is too short to observe a clear, reproducible phenotype attributable to defective XCI. In order to further deepen the role of XCI in the acquisition of a TE-like cell profile, RNA-FISH experiments could be repeated at later time points targeting *POLA1* and other X-linked genes known to undergo transcriptional silencing (*ATRX, MECP2*, *HUWE1*). Indeed, TE cells obtained with the described protocol could be maintained in culture until day 7 even though were shown to progressively upregulate cytotrophoblast and syncytiotrophoblast markers[27]. Moreover, TE cells obtained with this conversion protocol have shown to be competent for further commitment into the trophoblast lineage, generating TSCs, which can themselves be differentiated to syncytiotrophoblasts and extravillous cytotrophoblasts. Hence, it would be interesting to observe if and when *XIST*/XCI repression becomes problematic in this specification/commitment/differentiation scheme.

The identity of this cell line at day 5 was characterized by low expression of pluripotency and naïve hESC markers, while trophectoderm and TSC markers tended to be upregulated. The fact that some markers (*NR2F2*, *GATA3*, *GATA2*) but not others (*ENPEP*, *CDX2*) were upregulated in both DOX+ and DOX- conditions may circumscribe *day-5* cells within a specific temporal window around the late TE /early cytotrophoblast stage. Alternatively, the cells obtained with this procedure

might not correspond to proper TE cells, an issue that needs to be further assessed by (1) comparing the transcriptional profile of *day-5* TE with *ex vivo* and *in vitro* gene expression datasets and (2) testing the ability to produce trophoblast derivatives (as previously mentioned).

Importantly, hESCs were maintained in hypoxic conditions throughout the TE specification protocol. This aspect may be relevant since it was not specified in the *Guo et al.* reference publication. For sure, it would be useful to repeat the experiments in normoxia so to evaluate the impact of O_2 concentration during the process.

As for the correspondence between these observations and the dynamics occurring *in vivo*, it is important to remind that naïve hESCs are not perfectly reliable models for human embryonic development, since they are cells in a bidimensional artificial condition that forces them to remain into a physiologically transient state. These cells are particularly sensitive to manipulation and changes in the cell culture conditions. They were shown to undergo spontaneous differentiation after few tens of passages, limiting their prolonged maintenance in culture and potentially biasing the study of X chromosome dynamics in both the naïve and committed states. This is another reason why the assessment of XCI by RNA-FISH should be (1) extended to a collection of X-linked genes instead of one and (2) be coupled with immunostaining or FACS to more precisely determine the identity of tested cells.

However, even in their optimal state naïve hESCs cells show striking differences with what observable *in vivo*, most notably H3K27me3 and *XIST* coating being enriched in only one X chromosome. For all these reasons, new models should be introduced to test the aforementioned conclusions. For example, the implementation of new female hESC lines would likely remove cell line-specific biases. Alternatively, tridimensional cultures could level up the study to a context that is more similar to an actual embryo in terms of cellular organization and cell-to-cell interactions. Blastoids are structures resembling the blastocyst state at both the morphological and the transcriptional levels. They can be obtained *in vitro* starting from a culture of naïve hESCs[35]. Since they allow TE cells to assume their natural organization, they would be a rich source of information to confirm or disconfirm the results obtained so far. Moreover, their use in this context would likely shed light on (1) the involvement of XCI in subsequent steps of embryonic development that cannot be mimicked by 2D cultures, such as implantation; (2) the

impact of XCI impairment on both TE and ICM-like cells, in terms of cell state, differentiation capability and crosstalk.

In conclusion, XCI does not seem to be required for the commitment to an extraembryonic fate, even though it seems to be essential for the further differentiation and stability of trophectoderm derivatives. Further studies are needed to identify the critical time window during which XCI is essential for developmental progression and cell survival. These would likely have key implications on multiple levels, from the research field of assisted reproductive technology to the sphere of evolutionary biology.

I would like to express my profound gratitude to my supervisor and the host team: they gave me invaluable support and care, without which I would not have gone so far and enjoyed so much this scientific adventure.

This work is dedicated to my family, my best friends and *mon cœur* Francesco: the shiniest stars in my sky.

BIBLIOGRAPHY

1. Barr M, Bertram E. **A Morphological Distinction between Neurones of the Male and Female, and the Behaviour of the Nucleolar Satellite during Accelerated Nucleoprotein Synthesis**. *Nature*. 1949; 163: 676–7.

2. Lyon MF. **Gene Action in the X-chromosome of the Mouse (*Mus musculus* L.)**. *Nature*. 1961; 190: 372–3.

3. Shevchenko AI, Zakharova IS, Zakian SM. **The evolutionary pathway of X chromosome inactivation in mammals**. *Acta Naturae*. 2013; 5(17): 40–53.

4. Subrini J, Turner J. **Y chromosome functions in mammalian spermatogenesis**. *Elife*. 2021; 10:1–20.

5. Libert C, Dejager L, Pinheiro I. **The X chromosome in immune functions: When a chromosome makes the difference**. *Nature Reviews Immunology*. 2010; 10(8): 594–604.

6. Vallot C, Ouimette JF, Rougeulle C. **Establishment of X chromosome inactivation and epigenomic features of the inactive X depend on cellular contexts**. *BioEssays*. 2016;38(9):869–80.

7. Nguyen DK, Disteche CM. **Dosage compensation of the active X chromosome in mammals**. Nature Genetics. 2006 Jan; 38(1): 47-53.

8. Gilbert S. F. **Developmental Biology**. 9th edition. *Oxford University Press*. 2010. Chapter 8: 300-311; Chapter 17: 649-651.

9. Mihajlović AI, Bruce AW. **The first cell-fate decision of mouse preimplantation embryo development: Integrating cell position and polarity**. *Open Biology*. 2017; 7(11).

10. Martin GR. **Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells**. *Proceedings of the National Academy of Science of the United States of America*. 1981; 78(12 II): 7634–8.

11. Evans M J, Kaufman M H. **Establishment in culture of pluripotential cells from mouse embryos**. *Nature*. 1981; 292(July): 154–6.

12. Zhao W, Li Y, Zhang X. **Stemness-Related Markers in Cancer**. *Cancer Translational Medicine*. 2017; 3(3): 87-95.

13. Takahashi K, Yamanaka S. **Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors**. *Cell*. 2006; 126(4): 663–76.

14. Papatsenko D, Waghray A, Lemischka IR. **Feedback control of pluripotency in embryonic stem cells: Signaling, transcription and epigenetics**. *Stem Cell Research*. 2018; 29: 180–8.

15. Patrat C, Ouimette JF, Rougeulle C. **X chromosome inactivation in human development**. *Development.* 2020; 147(1).

16. Mutzel V, Schulz EG. **Dosage Sensing, Threshold Responses, and Epigenetic Memory: A Systems Biology Perspective on Random X-Chromosome Inactivation**. *BioEssays*. 2020; 42(4): 1–14.

17. Loda A, Heard E. **Xist RNA in action: Past, present, and future**. *PLOS Genetics*. 2019; 15(9): 1–17.

18. Gendrel AV, Heard E. **Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation**. *Annual Review of Cell and Developmental Biology*. 2014; 30: 561–80.

19. Robert-Finestra T, Tan BF, Mira-Bontenbal H, et al. **SPEN is required for Xist upregulation during initiation of X chromosome inactivation**. *Nature Communications*. 2021; 12(1): 1–13.

20. Pinheiro I, Heard E. **X chromosome inactivation: New players in the initiation of gene silencing**. *F1000Research*. 2017; 6(0).

21. Hemberger M. **The role of the X chromosome in mammalian extraembryonic development**. *Cytogenetic and Genome Research*. 2002; 99(1–4): 210–7.

22. Zhang LF, Huynh KD, Lee JT. **Perinucleolar Targeting of the Inactive X during S Phase: Evidence for a Role in the Maintenance of Silencing**. *Cell*. 2007; 129(4): 693–706.

23. Masui O, Corbel C, Nagao K, et al. **Polycomb repressive complexes 1 and 2 are each essential for maintenance of X inactivation in extra-embryonic lineages**. *Nature Cell Biology*. 2023; 25(1): 134–44.

24. Okamoto I, Nakamura T, Sasaki K, et al. **The X chromosome dosage compensation program during the development of cynomolgus monkeys**. *Science*. 2021; 374(6570).

25. Matthews KRW, Moralí D. **National human embryo and embryoid research policies: A survey of 22 top research-intensive countries**. *Regenerative Medicine*. 2020; 15(7): 1905–17.

26. Thomson JA. **Embryonic stem cell lines derived from human blastocysts**. *Science*. 1998; 282(5391): 1145–7.

27. Guo G, Stirparo GG, Strawbridge SE, et al. **Human naive epiblast cells possess unrestricted lineage potential**. *Cell Stem Cell*. 2021; 28(6): 1040-1056.

28. Rostovskaya M, Stirparo GG, Smith A. **Capacitation of human naïve pluripotent stem cells for multi-lineage differentiation**. *Development*. 2019; 146(7): 1–15.

29. Guo G, von Meyenn F, Rostovskaya M, et al. Correction: Epigenetic resetting of human pluripotency. *Development*. 2018 Apr 18; 145(8): dev166397. Erratum for: **Epigenetic resetting of human pluripotency**. *Development*. 2017 Aug 1; 144(15): 2748-2763.

30. Castel G, Meistermann D, Bretin B et al. **Induction of Human Trophoblast Stem Cells from Somatic Cells and Pluripotent Stem Cells**. *Cell Rep*orts. 2020; 33(8).

31. Krendl C, Shaposhnikov D, Rishko V, et al. **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 Science of the United States of America*. 2017; 114(45): E9579–88.

32. Schertzer MD, Thulson E, Braceros KCA, et al. **A piggyBac-based toolkit for inducible genome editing in mammalian cells**. *RNA*. 2019; 25(8): 1047–58.

33. Io S, Kabata M, Iemura Y, et al. **Capturing human trophoblast development with naive pluripotent stem cells in vitro**. *Cell Stem Cell*. 2021; 28(6): 1023-1039.

34. Frankenberg S, Shaw G, Freyer C. **Early cell lineage specification in a marsupial: A case for diverse mechanisms among mammals**. *Development*. 2013; 140(5): 965–75.

35. Yu L, Wei Y, Duan J, et al. **Blastocyst-like structures generated from human pluripotent stem cells**., *Nature*. 2021.; Vol.591: 620–626.