Das Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft (MDC)

DISSERTATION

Humane endogene Retroviren-gesteuerte Expression von schwangerschaftsspezifischem Glykoprotein 9 bei Plazentation und Prä-eklampsie

Human Endogenous Retroviruses-Directed Expression of Pregnancy-Specific Glycoprotein 9 in Placentation and Pre-eclampsia

zur Erlangung des akademischen Grades

Doctor of Philosophy (PhD)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Yuliang Qu

aus Shandong, China

Datum der Promotion: 25.11.2022

Table of contents

List of tables

List of figures

List of abbreviations

APA	Alternative polyadenylation
aPKC	atypical protein kinase C
Cas 9	CRISPR associated protein 9
cAMP	Cyclic adenosine monophosphate or cyclic AMP
CCC	Cytotrophoblast cell column
CEA	Carcinoembryonic antigen
CEACAMs	Carcinoembryonic antigen-related cell adhesion molecules
CNV	Copy number variation
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
dpf	days post fertilization
ECM	Extracellular matrix
EMT	Epithelial-to-mesenchymal transition
Env	Envelop glycoprotein
EO-PE	early onset pre-eclampsia
EPAC	Exchange protein directly activated by cAMP
ER	Endoplasm reticulum
ERVs	Endogenous retroviruses
ESCs	Embryonic stem cells
EVTs	Extravillous trophoblasts
eEVTs	endovascular EVTs
Gal -1	Galectin -1
GATA2	GATA-binding proteins 2
H3K4me1	Histone H3 lysine 4 monomethylation
H3K27ac	Histone H3 lysine 27 acetylation
HERVs	Human endogenous retroviruses
Hi-C	High throughput chromosome conformation capture
HIF	Hypoxia-inducible factor
HLA -G	Human leucocyte antigen G
HSPG	Heparan sulfate proteoglycan
hTSCs	human trophoblast stem cells

ICM	Inner cell mass
iEVTs	interstitial EVTs
KLF6	Krüppel-like transcription factor 6
LHCGR	LH chorionic gonadotropin receptor
LINE1	Long interspersed nuclear element -1
mRNA	messenger ribonucleic acid
LO-PE	late onset pre-eclampsia
LTR	Long terminal repeat
PAS	Polyadenylation signal
PLAC8	Placenta-specific protein 8
PLGF	Placental growth factor
PSGs	Pregnancy-specific glycoproteins
PE	pre-eclampsia
PKA	Protein kinase A
RARE	Retinoic acid responsive element
sENG	soluble Endoglin
SCTs	Syncytiotrophoblasts
sFLT1	soluble fms-like tyrosine kinase 1
SINE	Short interspersed nuclear element
SNP	Single-nucleotide polymorphism
Sp1	Specificity protein 1
TAD	Topologically associating domain
TE	Trophectoderm
TEs	Transposable elements
TEAD4	TEA-domain family member 4
TFAP2A	Transcription factor AP-2 gamma 2A
TGF-β	Transforming growth factor β
TSCs	Trophoblast stem cells
TSFs	Trophoblast specification factors
UPR	Unfolded protein response
VCTs	Villous cytotrophoblasts
VEGF	Vascular endothelial growth factor
YAP1	Yes-associated protein 1

Abstract

The pregnancy-specific glycoprotein (PSG) has been recognized as one of the most abundant polypeptide hormones in maternal blood during human pregnancy [1,2]. The *PSG* genes (*PSG1-PSG11*) are clustered at human chromosome 19q13 and predominantly expressed in placental trophoblasts [1]. PSGs are known to contribute to several important processes in pregnancy such as angiogenesis and immunomodulation [3,4], whereas abnormal levels of PSGs in maternal blood have been detected in pregnancy disorders including pre-eclampsia (PE) [5,6]. Despite the physiological and pathological implications, neither the regulation nor biological function of PSG expression in human trophoblasts has been extensively explored. Given the emerging roles of endogenous retroviruses (ERVs) in placental development [7,8], we interrogated the regulation of human ERVs (HERVs) in PSG expression. By mining the online epigenetic data and experimental validation, we identified two elements, LTR8 and MER65-int, which could control PSG9 expression at different levels. The LTR8 promotes PSG9 transcription as an enhancer bound by the trophoblast specification factors, GATA2 and GATA3, whereas the MER65-int is involved in the mRNA polyadenylation of PSG9, producing the secretory isoform. Importantly, the loss-of-function of either LTR8 or PSG9 results in defective trophoblast syncytialization to variable extents, indicating the HERVs-directed PSG9 is essential for syncytiotrophoblast differentiation. On the other hand, we examined PSG expression in pregnancy complications and detected that mRNA levels of the secretory PSG9 isoform are specially upregulated in trophoblasts of early onset PE (EO-PE), supporting that increased PSG9 in maternal blood could be a potentially valuable biomarker of this disease. Altogether, the project provides insights into the versatile roles of HERVs in gene regulation and sheds light on the biological function of HERVs-directed PSG9 in trophoblast differentiation.

Zusammenfassung

Das schwangerschaftsspezifische Glykoprotein (PSG) wurde als eines der am häufigsten vorkommenden Polypeptidhormone im mütterlichen Blut, während der menschlichen Schwangerschaft, erkannt [1,2]. Die PSG-Gene (*PSG1-PSG11*) befinden sich auf dem menschlichen Chromosom 19q13 eng geclustert und werden vorwiegend in plazentaren Trophoblasten exprimiert [1]. Es ist bekannt, dass PSGs an mehreren wichtigen Prozessen in der Schwangerschaft beteiligt sind, wie z.B. der Angiogenese und der Immunmodulation [3,4], wohingegen abnormale Konzentrationen von PSGs im mütterlichen Blut bei Schwangerschaftsstörungen einschließlich Präeklampsie (PE) nachgewiesen wurden [5,6]. Trotz der physiologischen und pathologischen Implikationen sind weder die Genregulation noch die biologische Funktion der PSG-Expression in menschlichen Trophoblasten umfassend erforscht. Angesichts der sich abzeichnenden Rolle von endogenen Retroviren (ERVs) in der Plazenta-Entwicklung haben wir die Regulation durch humanen ERVs (HERVs) in der PSG-Expression untersucht [7,8]. Mit Hilfe der Auswertung von epigenetischen Online-Daten und experimentelle Validierung identifizierten wir zwei Elemente, LTR8 und MER65-int, die PSG9-Expression auf verschiedenen Ebenen kontrollieren. LTR8 fördert die PSG9-Transkription als Enhancer, der von den Trophoblasten-Spezifikationsfaktoren GATA2 und GATA3 gebunden wird, während MER65-int an der mRNA-Polyadenylierung von PSG9 beteiligt ist, wodurch die sekretorische Isoform entsteht. Wichtig ist, dass der Funktionsverlust von entweder LTR8 oder PSG9 zu einer defekten Trophoblasten-Synzytialisierung in unterschiedlichem Ausmaß führt, was darauf hindeutet, dass das von HERVs gesteuerte PSG9 für die Synzytiotrophoblastendifferenzierung essentiell ist. Andererseits untersuchten wir die PSG-Expression bei Schwangerschaftskomplikationen und stellten fest, dass der mRNA-Spiegel der sekretorischen PSG9-Isoform, speziell in Trophoblasten der früh einsetzenden PE (EO-PE), hochreguliert ist. Das spricht dafür, dass eine erhöhte PSG9-Konzentration im mütterlichen Blut, ein potentiell wertvoller Biomarker für diese Erkrankung sein könnte. Insgesamt veranschaulicht das Projekt die vielseitigen Rollen von HERVs in der Genregulation und wirft Licht auf die biologische Funktion von HERV-gesteuertem PSG9 in der Trophoblastendifferenzierung.

1. Introduction

1.1 Human placental development and pre-eclampsia (PE)

The human placenta is a temporary organ of fetus during pregnancy. It plays multiple roles in maintaining fetal and maternal health during gestation and even has a lifelong influence on their future wellbeing [9-11]. The placental dysfunction or impaired placentation has been implicated in pregnancy complications, including pre-eclampsia (PE), fetal growth restriction (FGR), recurrent miscarriage and still-birth [12,13]. Despite the significances, our understanding of human placental development remains limited, largely due to the lack of ideal experimental models. Recent advances in developing novel systems and techniques, such as human trophoblast stem cell culture, 3D trophoblast organoids and single-cell multiomics, have cast placental research back into the spotlight.

1.1.1 An anatomical overview of human placental development

Embryogenesis occurs upon the effective combination of oocyte and sperm. Thereafter, the single-celled zygote undergoes several rounds of cleavage divisions, producing a mass of totipotent cells, termed morula. With the specification of the first definitive cell lineages, the human blastocyst stage begins at around 5 days post - fertilization (dpf). The human placenta originates from the outer trophectoderm (TE) of blastocyst, whereas the embryo proper is differentiated from the inner cell mass (ICM) (Fig. 1A).

Upon implantation at around 6 - 7 dpf, the polar TE which is adjacent to ICM has been proposed to attach to the maternal endometrium and to fuse into the multinuclear primary syncytium [9,10]. Then the primary syncytium breaks through the luminal epithelium and invades into the decidualizing endometrium to anchor the blastocyst to the maternal uterine wall [9,10,14]. The mononuclear cytotrophoblasts (VCTs) at the inner layer further rapidly proliferate and penetrate through the primary syncytium to form the primary villi, whereby the placental villus tree is gradually shaped through more proliferation and branching [9,10].

The proliferative cytotrophoblasts extensively expand and eventually encapsulate the conceptus as a continuous cytotrophoblast shell. In the days that follow, the secondary villi are formed with penetration of extraembryonic mesenchymal cells into

the primary villi [10,15]. At around 18 dpf, upon the generation of fetal blood capillaries within the villus core, the secondary villi are further transformed into the tertiary villi [10,15] (Fig. 1A). The villous trees continue to rapidly expand through the progressive branching from the chorionic plate. Meanwhile, the invasive extravillous trophoblasts (EVTs) are differentiated from the cytotrophoblast shell at the maternal-fetal interface. In this way, this basic structure of the human placenta is formed by around the fourth week of gestation, while the mature hemochorial placenta is gradually established by the end of the first trimester when the perfusion with maternal blood is enabled with the remodeling of the uterine spiral arteries by EVTs [9,10,15,16] (Fig. 1A and B).

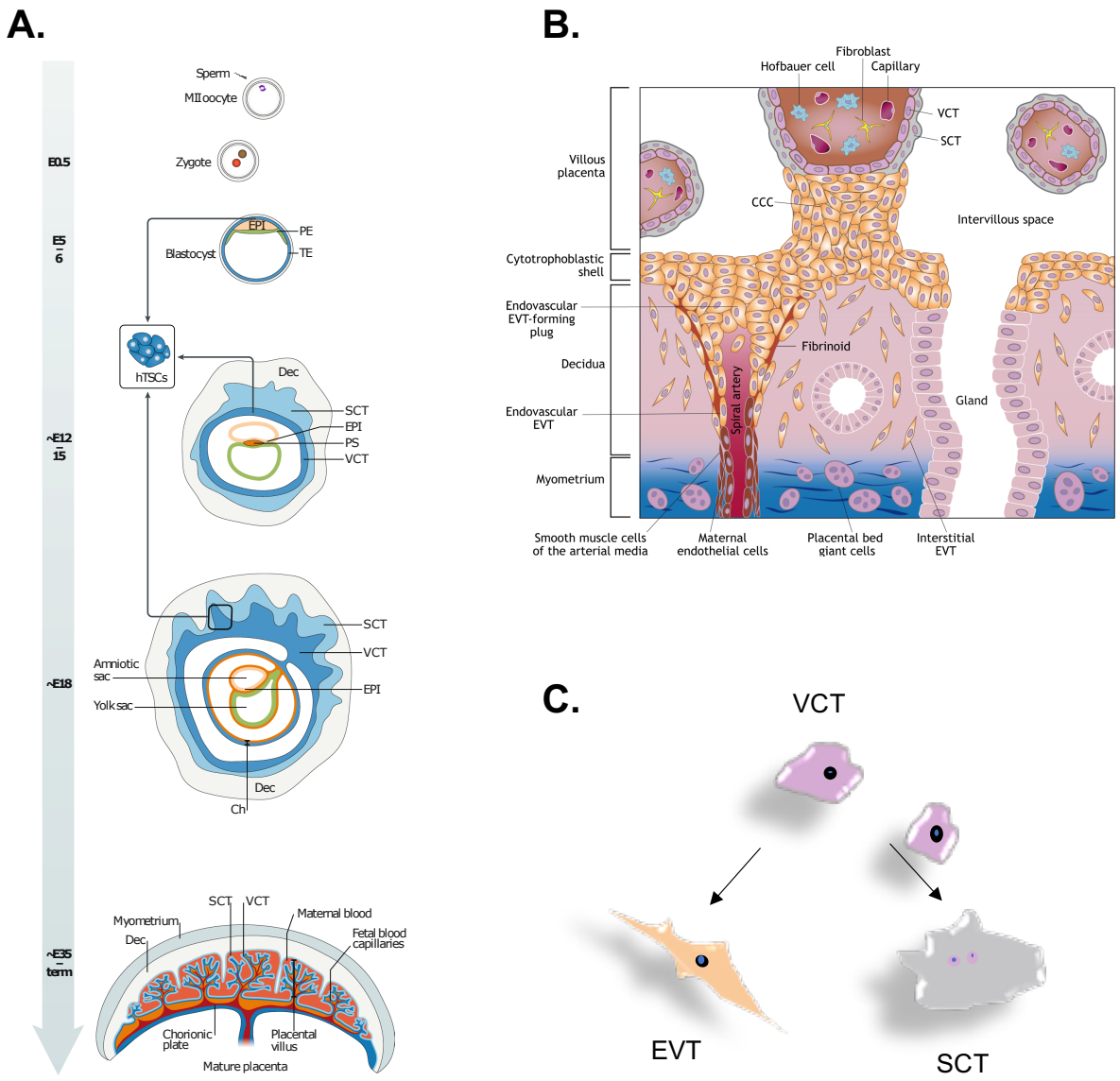

Fig. 1 An anatomical overview of human placentation. (**A**) The early stages of human placental development over embryonic days (E). Human TSCs could be derived from the trophoblast cells of either the blastocyst or the early first-trimester placenta. Adapted from [9] with the license provided by

Springer Nature and Copyright Clearance Center (**B**) An overview of major trophoblast subtypes at the maternal-fetal interface in the early first trimester, adapted from [10] with the license provided by the Company of Biologists. (**C**) Diagram depicting the differentiation of VCT into EVT and SCT trophoblast cells in the first trimester placenta. EPI, epiblast; PE, primitive endoderm; TE, trophectoderm; Dec, decidua; PS, primitive streak; Ch, chorion; CCC, cytotrophoblast cell column; VCT, villous cytotrophoblast; SCT, syncytiotrophoblast; EVT, extravillous trophoblast.

1.1.2 Molecular mechanism of human trophoblast differentiation

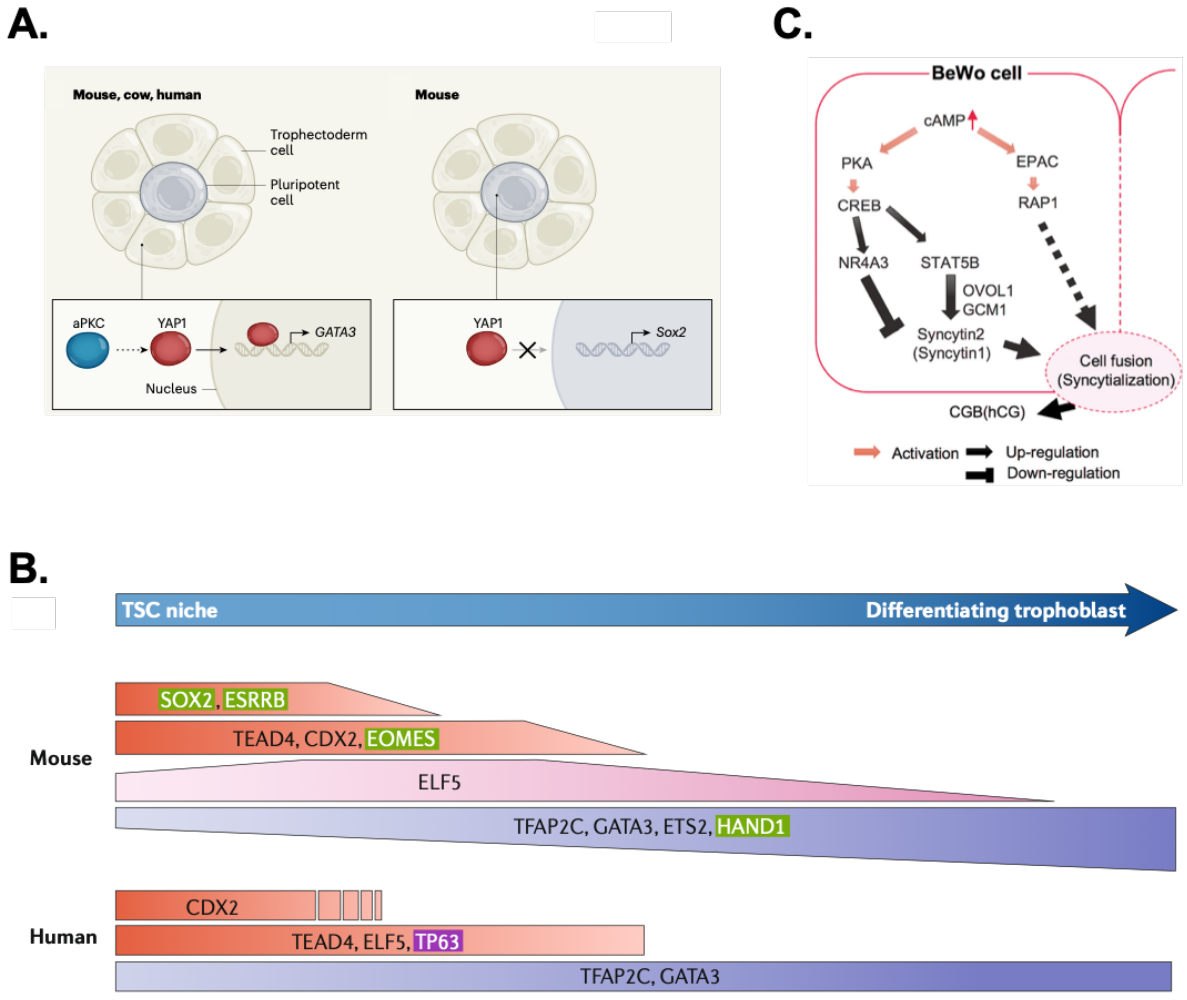

Fig. 2 Transcriptional controls of human trophoblast differentiation. (**A**) A conserved molecular cascade, 'aPKC-YAP1/TEAD4-GATA3', could initiate the TE specification at the morula stage across mouse, cow and human. In mouse morula, the inhibition of YAP1 movement into the nucleus leads to SOX2 expression in inner cells, whereas the mechanisms governing the establishment of pluripotency in cow and human embryos remain unclear [17]. (**B**) Diagram depicting the relative abundance of key transcription factors during mouse and human VCT differentiation. In human early placenta, the VCTs have trophoblast stem or progenitor cell features. The factors highlighted in green are vital for mouse

mTSCs, while they seem to be absent in human TSCs [9]. TP63, highlighted in purple, is expressed in human VCTs. (**C**) The canonical 'cAMP-PKA/EPAC' signaling for trophoblast syncytialization. **(A) and (B)** adapted from [17] and [9], respectively, with the licenses provided by Springer Nature and Copyright Clearance Center. (**C**) adapted from [18] with the license provided by John Wiley and Sons and Copyright Clearance Center.

1.1.2.1 Specification of human trophoblast lineage

Our knowledge of the molecular mechanism underpinning trophoblast differentiation has been obtained largely based on the mouse model. The first cell-fate decision initiates at the morula stage, leading to specification of the embryonic and trophoblast lineages by the blastocyst stage. A plethora of factors driving early lineage commitment have been identified. Notably, a recent report has revealed a conserved molecular cascade, which could initiate the TE programme across cow, mouse and human [19]. With the atypical protein kinase C (aPKC) at the contact-free domain, the outer cells of morula embryo are polarized along the apical–basal axis. Then the accumulating aPKC sequesters the Angiomotin (AMOT) at the apical domain to keep the Hippo signalling inactive, leading to translocation of Yes-associated protein 1 (YAP1) to the nucleus. The nuclear YAP1 and its cofactor TEA-domain family member 4 (TEAD4) further promote the restricted expression of TE-defined factors, such as GATA3, to initiate TE specification [17,19] (Fig. 2A).

Despite the striking conservation, differences in the expression pattern of lineage specification factors still remain. For instance, in the mouse morula, the expression of SOX2, an early marker of ICM pluripotency, is specially restricted to the inner cells through inhibiting the movement of YAP1 protein into the nucleus. By contrast, SOX2 is broadly expressed in cow and human embryos up to the early blastocyst stage [17,19]. In addition, the 'YAP1/TEAD4' complex - activated CDX2 has been long considered as another key player in mouse TE specification by mutually antagonizing with OCT4 expression prior to the blastocyst formation [9,20]. However, CDX2 seems to be later expressed at cow or human blastocyst stage [17,21,22].

1.1.2.2 Differentiation of villous cytotrophoblasts (VCTs)

The molecular events controlling the primary villi formation at human early implantation stage (around 7dpf) remain in the 'black box', and the studies of human

VCT specification have been largely descriptive [10,11]. The VCTs of human early placenta retain trophoblast stem or progenitor cell characteristics. Similar to the case in mice, TEAD4 and CDX2 could act in parallel with multiple other transcription factors to promote VCT differentiation, including GATA-binding proteins 3 (GATA3) and Transcription factor AP-2 gamma 2A (TFAP2C) [9,23-25]. Interestingly, ESRRB, SOX2 and EOMES which are essential for mouse TSC maintenance seem to be absent or play a less prominent role in human TSCs [9,26-29] (Fig. 2B). Of note, CDX2 remains highly expressed in VCTs of the early first-trimester placenta and is rapidly downregulated thereafter, whereas the expression of GATA3 and TFAP2C gradually increases over this process [24,26,30] (Fig. 2B). In addition, ELF5 and the tumour protein TP63 are widely expressed in human VCTs [26,31,32]. With the CDX2 expression, the VCTs expressing either ELF5 or TP63 have been thought to be the best candidate cells with stem cell potential in human first-trimester placenta [9,30,31].

1.1.2.3 Differentiation of syncytiotrophoblasts (SCTs)

The SCTs are derived from the underlaying VCTs through the cell-cell fusion procedure, termed trophoblast syncytialization. Current insights into SCT differentiation have been gained based on the studies in mice and in *vitro* cell-based models. Particularly, the canonical signaling pathway for trophoblast syncytialization has been well characterized using the Forskolin-induced BeWo cell fusion system. Upon treatment of Forskolin, the intracellular cyclic AMP (cAMP) activates two downstream signaling molecules, protein kinase A (PKA) and exchange protein directly activated by cAMP (EPAC) [18]. The 'cAMP-PKA-CREB' signaling could upregulate STAT5B to promote Syncytin-2 expression which is mediated by the glial cell missing transcription factor 1 (GCM1) and OVO-like 1 (OVOL1) [18,33-36]. Meanwhile, the same signaling has been thought to upregulate NR4A3 and to inhibit cell fusion, suggesting a role in coordinating the degree of trophoblast syncytialization [18] (Fig. 2C). On the other hand, the 'cAMP-EPAC' signaling may induce syncytialization through activating RAP1 for GCM1 phosphorylation [18,37-39] (Fig. 2C). Syncytin protein, encoded by the endogenous retroviral *env* genes, has long been recognized a critical player in SCT formation [7,40]. In human placenta, Syncytin-1 and Syncytin-2 mediate the fusion of the VCT into SCT with binding to the receptor ASCT2 and MFSD2, respectively (see details below) [41,42]. Upon effective fusion, the functional syncytia can produce

and secrete multiple pregnancy hormones, such as human chorionic gonadotropin (hCG) and progesterone [18,43]. Interestingly, the hCG has been proposed to could regulate SCT formation in a 'positive-feedback loop', where the soluble hCG binds to the LH chorionic gonadotropin receptor (LHCGR) and further induce the cAMP signaling for syncytialization [10,44,45].

In addition to the classical signaling, novel molecular events behind SCT differentiation have been recently revealed. Notably, increasing evidence has shed light on the cell cycle exit as a key prerequisite for proper syncytialization [11]. In human placenta, the complex of the cell-cycle inhibitor p21 and the transcription factor GCM1 promote restricted expression of Syncytin-2 in G0 phase to guarantee appropriate trophoblast fusion. Indeed, overexpression of Syncytin-2 in cycling cells results in functionally compromised syncytia [46]. Besides, it has been recently identified that TMEM16F, a Ca2+-activated phospholipid scramblase (CaPLSase), plays an essential role in trophoblast fusion [47]. By contrast, the 'YAP1/TEAD4' complex, a crucial module for TE specification at early embryo stage, seems to negatively regulate SCT differentiation from VCTs [48]. Other factors produced under pathological conditions could also have a role in SCT formation. For instance, during viral infection, the type I interferon (IFN)-induced transmembrane proteins (IFITMs) are known to protect uninfected cells from viral entry [49]. However, the induced IFITMs impair Syncytin-mediated cell fusion and thus affect fetal development, providing a molecular explanation for abnormal placentation in IFN-mediated pregnancy disorders [50].

1.1.2.4 Differentiation of extrovillus trophoblasts (EVTs)

In human first-trimester placenta, the EVTs originate from the cytotrophoblast cell columns (CCCs) at the tips of anchoring villi which are in direct contact with the maternal decidua [10,15]. The EVTs can be further divided into two main populations: interstitial EVTs (iEVTs) and endovascular EVTs (eEVTs). The iEVTs invade the decidual stroma as far as the inner third of the myometrium, where they fuse and form the multinucleated trophoblast giant cells [15,51]. On the other hand, the eEVTs, which embed in the maternal spiral arteries, aggregate as a plug to prevent blood from flowing into the intervillous space until the end of the first trimester when the mature haemochorial placenta is fully established [10,16,52].

The molecular mechanism controlling EVT differentiation remains obscure. It has been proposed that the transition of EVTs from VCTs initiates under low oxygen tension, where the hypoxia-inducible factor (HIF) complex-mediated signalling seems to promote this process [10,53]. Multiple transcription factors are rapidly upregulated during the transition, including ASCL2, TEAD2, TCF4, HIF1A and GCM1, whereas TP63, a VCT-specific transcription factor, is decreased [53]. In addition, Notch1 has been reported as a key regulator in EVT specification by inducing markers of EVT progenitors but repressing genes controlling VCT self-renewal [54]. Furthermore, the EVT differentiation has been recognized to undergo the epithelial-to-mesenchymal transition (EMT) with downregulation of the epithelial cell markers such as Epithelial-cadherin (E-cadherin) and zonula occludens-1 (ZO-1) [55-58], and upregulation of the invasive cell markers such as Neuronal cadherin (N-cadherin) [59,60]. Of note, the 'integrin switching' is induced over EVT differentiation from the laminin-binding integrin a6β4 to the fibronectin-binding integrins a1β1 and a1β5 [61,62]. Particularly, the upregulation of the major histocompatibility complex class I (MHC I) molecules, human leucocyte antigen G (HLA-G) and HLA-C, has been regarded as a defining marker of EVT differentiation [10,63,64]. Nevertheless, the molecular events behind the specification of distinct EVT subtypes need to be clarified. The Placenta-specific protein 8 (PLAC8) has recently been identified as a novel marker of iEVT cells [65], whereas a polysialylated form of neural cell-adhesion molecule (NCAM) is expressed in eEVTs [66].

1.1.2.5 Advances in models and techniques for human placental research

Due to the lack of ideal models which can recapitulate human placentation, investigation of human trophoblast differentiation has long been impeded. Until recently, multiple breakthroughs in human placenta research have been achieved. In 2018, the 'bona fide human trophoblast stem cells (hTSCs)' were successfully established from the CTBs derived from the blastocyst or first-trimester placenta [67]. The resulting hTSCs are genetically stable and fulfil the most widely accepted criteria of first-trimester trophoblast [67,68]. Importantly, these hTSCs show the transcriptome and methylome similar to those of primary CTBs and are able to differentiate into the SCT- and EVT-like cells [67]. Besides those generated from hESCs with the BAP treatment [69,70], the hTSCs have been further induced from pluripotent stem cells

(PSCs) or even somatic cells through direct reprogramming [71-73], which could not only circumvent accessing the human early embryo and first-trimester placenta, but also hold promise to establish the patient-specific hTSCs.

Furthermore, the 3D culture of human trophoblasts, termed trophoblast organoids, has been established [74,75]. Human trophoblast organoids can recapitulate not only features of trophoblasts but also the complex branching morphology of early placental villi [75]. In spite of the potential ethic debates, further efforts are ongoing to model early embryo development in *vitro* by co-culturing hTSCs or trophoblast organoids with the endometrial glands, endometrial stroma and hESCs, opening avenues to explore the crosstalk between placenta and other tissues [9,76,77]. Collectively, these transformative advances in the 2D or 3D culture of human trophoblasts would provide powerful tools to study placental development and related disease.

On the other hand, progression in single-cell transcriptomics and even multi-omics technologies has paved the way to decipher human placentation at unprecedented resolution [11]. Thanks to this, novel trophoblast subtypes and signature genes have been comprehensively appreciated [78]. Of note, Vento-Tormo *et al*. profiled the transcriptomes of about 70,000 single cells from human placentas with matched maternal blood and decidual cells, picturing the first comprehensive single-cell transcriptomic atlas of the maternal–fetal interface in the first trimester [79]. In addition, integrating with the plasma-circulating RNA analysis, the single-cell transcriptomics have provided more insights into placental cellular dysfunction in early pre-eclamptic patients [80]. As a non-invasive approach, this integrative analysis also holds promise to elucidate cellular dynamics in other complex biological or pathological processes.

1.1.3 Human placentation and pre-eclampsia (PE)

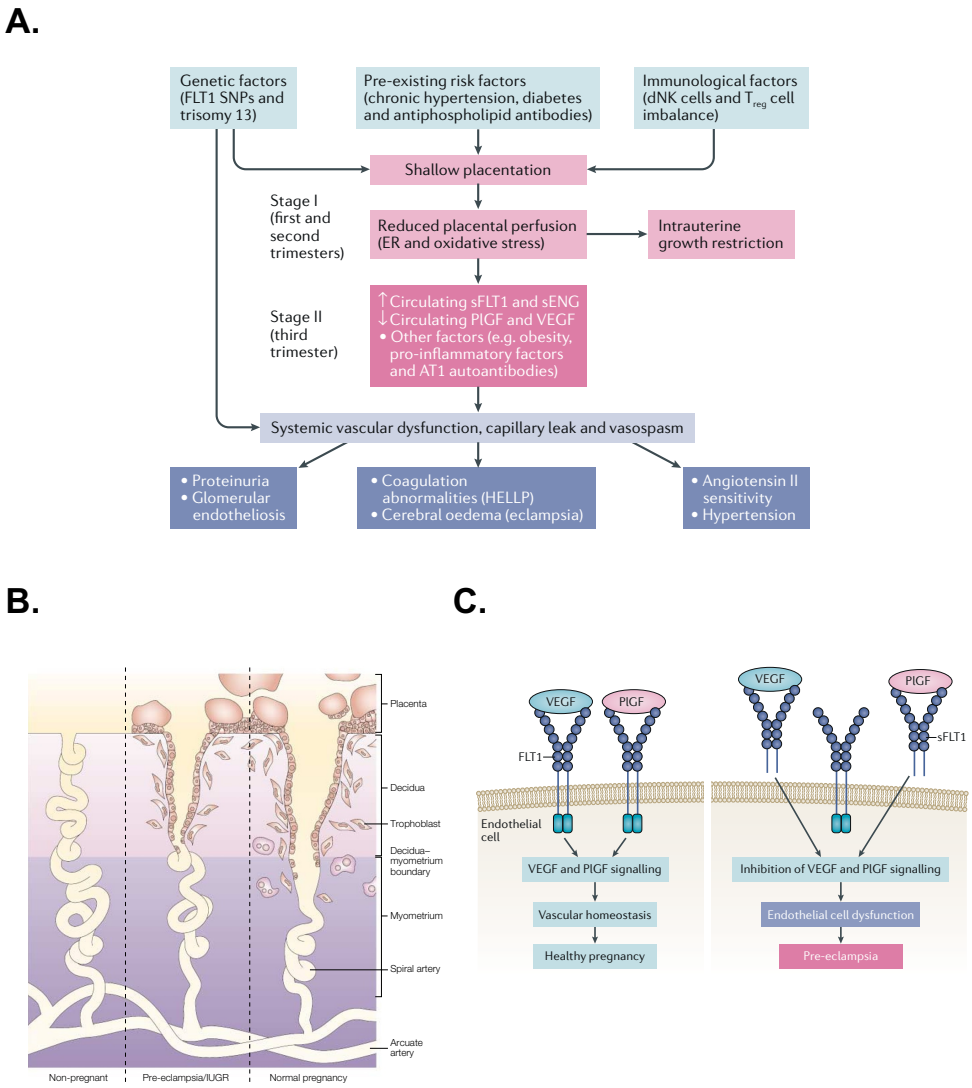

Fig. 3 The pathogenesis of PE. (**A**) The proposed 'two-stage' model of PE. Genetic factors, maternal pre-existing risk factors and immunological factors could cause defective trophoblast invasion, leading to a stress response in placenta (stage I); placental stresses further result in maternal endothelium dysfunction, systemic inflammatory response and clinical symptoms of PE (stage II) [12]. (**B**) Diagram depicting that defective trophoblast invasion into maternal decidua causes inadequate transformation of the spiral arteries in PE, leading to the ischemic placenta [81]. (**C**) Diagram depicting that excess circulating sFLT1 binds to the pro-angiogenic factors, VEGF and PLGF, inhibiting the FLT1 receptor-directed signaling for vascular homeostasis [12]. SNP, single-nucleotide polymorphism; dNK, decidual natural killer; T_{reg}, regulatory T cell; ER, endoplasm reticulum; sFLT1, soluble fms-like tyrosine kinase 1; sENG, soluble endoglin; PLGF, placental growth factor; VEGF, vascular endothelial growth factor; AT1, angiotensin II type I receptor; HELLP, haemolysis, elevated liver enzymes and low platelet count. Adapted from [12,81] with the licenses provided by Springer Nature and Copyright Clearance Center.

The functional unit of human mature placenta is the villous, which contains three major trophoblast subpopulations: VCTs, SCTs and EVTs. As described above, VCTs give rise to the multinucleated SCTs through trophoblast syncytialization (Fig. 1C). The SCTs form a protective cover lining the villous tree, where multiple steroid and polypeptide hormones are synthesized and secreted into maternal circulation for a successful pregnancy. Besides, the syncytium layer serves as the main site of gas and nutrient exchange at the maternal-fetal interface. In agreement with this, it has recently been revealed that human SCT formation is associated with enhanced macropinocytosis, which is proposed to facilitate nutrient uptake from the maternal environment, particularly under nutrient deprivation, to guarantee the fetal development [82]. On the other hand, the EVTs derived from the VCT population at the tips of anchoring villi are further separated into iEVTs and eEVTs. The iEVTs properly migrate through the maternal decidua towards the spiral arteries, whereas the eEVTs remodel the maternal spiral arteries to control the maternal blood flowing into the intervillous space over gestation [16,51]. Proper placentation is crucial for fetal growth and development. Indeed, a systematic phenotyping of mouse gene knockout has recently identified that 68% of 103 embryonic lethal lines display placental defects, emphasizing the central role of placentation in embryo development [83]. In human, abnormal placental development or placentation has been implicated with pregnancy complications, including PE, FGR, recurrent miscarriage and still-birth [12,13].

PE is a leading cause of both fetal and maternal morbidity and mortality during pregnancy, affecting around 5% of pregnancies worldwide [12,13]. Based on the current guidelines from the International Society for the Study of Hypertension in Pregnancy (ISSHP), PE is defined as the *de novo* hypertension after 20 weeks' gestation accompanied by one or more of the following new-onset conditions: proteinuria, other maternal organ dysfunction and/or uteroplacental dysfunction [84]. Despite its heterogeneity, PE is generally classified into the early onset (EO-PE) or late onset (LO-PE) up to whether the clinical symptoms manifest before or after 34 weeks' gestation, respectively. PE seems to be a disorder that only occurs in pregnant women and certain primates [85]. With the lake of a suitable model that recapitulates all aspects of the disease, our understanding of PE pathogenesis remains rather limited [86]. EO-PE has been widely acknowledged to originate from the defective placentation,

whereas LO-PE may be related to interactions between the placental aging and the maternal genetic predisposition to cardiovascular and metabolic disease [13].

In the classical two-stage model (Fig. 3A), EO-PE is associated with the shallow trophoblast invasion and incomplete remodelling of maternal spiral arteries during early pregnancy [87-89], leading to a spectrum of stress responses in the placenta, such as increased oxidative stress [90,91], activation of the unfolded protein response (UPR) [92] and excess particles shed from the syncytium layer into maternal circulation [93,94] (stage I). These placental stresses could further result in dysfunction of maternal endothelial cells [95,96], a systemic inflammatory response and the clinical symptoms of PE (stage II) [12,13] (Fig. 3A). Numerous placental factors have been proposed to trigger this process. Of note, the imbalance between the soluble fms-like tyrosine kinase 1 (sFLT-1) and its cell membrane-bound isoform FLT-1 (VEGFR1) plays a pathological role in the maternal endothelial dysfunction [97]. Excess soluble sFLT-1, probably released from the syncytial particles shed into the maternal circulation [98], antagonizes with the FLT-1 receptor to bind to the local pro-angiogenic factors, including vascular endothelial growth factor (VEGF) and placental growth factor (PLGF), decreasing the signalling pathway for endothelial cell function [96,99,100]. Indeed, sFLT-1 overexpression can lead to PE-like symptoms in rodent models [101-103], while the higher levels of sFLT-1 in maternal blood correlate with more severities of the disease in human [97,104]. Thus, the increased ratio of sFLT-1/PLGF has been well recognized as a reliable biomarker of PE [97,105]. In addition, another anti-angiogenic factor, soluble endoglin (sENG), has been identified to be increased in PE [106,107]. In rodents, co-expression and sFLT-1 and sENG can lead to more severe PE-like symptoms than that induced by either protein alone [107], suggesting a potentially synergistic action of the two anti-angiogenic factors in PE pathogenesis. Despite the promising progresses, the only available method to cure PE has been the placenta delivery, leading to the premature fetus and even long-term risks of maternal and fetal health [12,13]. Further efforts are definitely needed to effectively predict and prevent this disease.

1.2 Human pregnancy-specific glycoproteins (PSGs)

The pregnancy-specific glycoprotein (PSG) subgroup and the carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) constitute the carcinoembryonic antigen (CEA) family, a member of the immunoglobulin superfamily (IgSF) [1,108]. There are twelve CEACAM and ten PSG protein-coding genes identified in human genome [108,109]. CEACAMs are broadly expressed in numerous cell types, including epithelial, endothelial, and immune cells. They can be secreted, glycosylphosphatidyl inositol (GPI)-anchored or transmembrane proteins with either a short or a long cytoplasmic tail [110]. CEACAMs are well known to play diverse physiological and pathological roles, such as in cell-cell adhesion, immune response and carcinogenesis [111-113]. By contrast, PSGs have long been thought to be predominantly expressed in placental SCTs and then secreted into maternal circulation [1,2]. Although the clinical involvement of PSGs in human pregnancy was documented in as early as 1970s [114], our knowledge of their expression and biological function has been limited. With the lack of a specific antibody, functional delineation of a given PSG member has been further compounded. Recent advances in genome annotation, single-cell transcriptomics and CRISPR-based editing tools have provided unprecedented opportunities for PSG research.

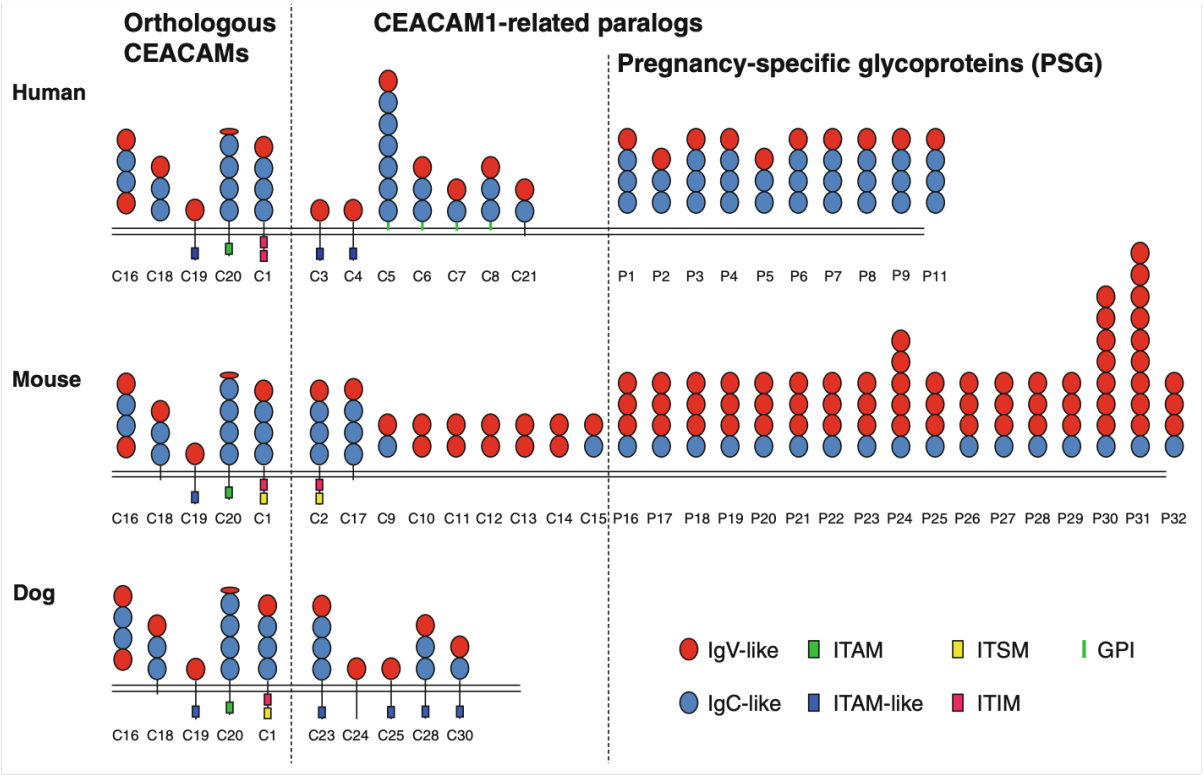

Fig. 4 The structural organization of mammalian CEA family members. The CEACAMs could be secreted, GPI-anchored or transmembrane proteins with either a short or a long cytoplasmic tail. Isoforms with the long cytoplasmic tail are mainly expressed in immune cells and provide inhibitory signals via ITIM and ITSM, while PSGs are predominantly expressed in trophoblasts and secreted into maternal circulation during pregnancy. Adapted from [115] with the license provided by Springer Nature and Copyright Clearance Center. IgV-like, Ig variable-like domain; IgC-like, Ig constant-like domain; ITIM, immunoreceptor tyrosine-based inhibitory motif; ITSM, immunoreceptor tyrosine-based switch motif; GPI, glycosylphosphatidyl inositol.

1.2.1 Human *PSG* gene and protein structure

Genes of the human *CEA* family are clustered on chromosome 19q13.2-19q13.31, where the ten protein-coding *PSG* members, *PSG1-PSG9* and *PSG11*, (*PSG10* has been recognized a transcribed pseudogene) are arrayed within a genomic region of 550 kilobase (kb) (GRCh38/hg38). The *CEA* gene family has been thought one of the most rapidly evolving gene families in the human genome [108,111,115-117]. Of note, the human *PSG* genes seem to be enriched with copy number variation (CNV) and nonsynonymous single nucleotide polymorphism (SNP), and exhibit a relatively high frequency of gene conversion between different *PSG* members compared to the chromosome average [1,118,119]. In consistent with this, a CNV deletion in human *PSG11* genomic region (*PSG9*, based on GRCh38/hg38) has been identified to be enriched in pre-eclamptic women (5/169 cases and 1/114 controls) [120]. Despite the association, whether the genetic variation can disturb expression of *PSG9* and further cause a phenotype is to be explored.

It has been proposed that PSGs may evolve independently in different mammalian species from a common founder gene, which likely leads to considerable variance in protein structure [116,121,122]. Human PSG protein is basically composed of a signal peptide, one N-terminal Ig variable-like domain (N domain), followed by two/three Ig constant-like domains (Ig C2-like domains of type A or B) and a relatively hydrophilic C-terminal domain (www.carcinoembryonic-antigen.de) (Fig. 4). Besides, there are a number of splicing and polyadenylation variants that differ especially in the carboxy terminus (C-terminus). Of note, notwithstanding the majority of PSGs as soluble hormones, one isoform with a long hydrophobic C-terminal domain (81 amino acid), termed PSG11W, was identified as an intracellular protein, which is retained in

endoplasmic reticulum (ER) [123-125]. Nevertheless, the regulation and function of this unique PSG isoform remains unexplored.

On the other hand, PSGs are highly glycosylated with carbohydrate content from 21 to 32% [2]. It has been recently reported that the glycan repertoire of human PSG1 is dominated by the multiantennary complex glycans with high levels of sialylation and partial core-fucosylation [126]. The glycans can mediate the interaction of PSG1 with Galectin-1 (Gal-1), a multifunctional protein during pregnancy, likely facilitating the stabilization of Gal-1 in the extracellular space [126].

1.2.2 Regulation of human PSG expression

PSGs are predominantly expressed in placenta and then secreted into maternal blood. With the increased levels in maternal circulation over gestation, PSGs have been regarded as one of the most abundant trophoblastic proteins [127-129]. Despite the abundance in bulk, the expression pattern of a defined PSG member and further the underlying regulatory mechanism have been rather obscure. It has been reported that promoters of *PSG* genes seem to be highly homologous with the lack of the classical TATA-box and pyrimidine-rich initiator element [130,131]. Nevertheless, several putative *cis*-elements in the proximal promoter region (around 600 bp upstream of translational start site) have been identified to control *PSG* transcription in trophoblasts [132]. For instance, the ubiquitous specificity protein 1 (Sp1) and the Krüppel-like transcription factor 4 (KLF4) could synergistically activate *PSG5* transcription through binding the FP1 *cis*-element and the Core Promoter Element (CPE: CCCCACCC), respectively [133,134]. Interestingly, the CPE also partially overlaps a putative Retinoic Acid Responsive Element (RARE) [132,135], suggesting a potentially regulatory role of retinoids signalling in *PSG* expression. Indeed, upon treatment with 9-*cis* retinoic acid (9cRA), *PSG5* transcription can be induced, which might be mediated by the interaction of Retinoid X receptor alpha (RXRa) with the RARE motif [135]. In agreement with this, retinoids and their receptors have been implicated in human placentation [136,137]. Besides the RARE motif, the putative binding site of GA binding protein (GABP), a Ets-family transcription factor, seems to be required for *PSG3* activation during SCT differentiation from the primary VCT in *vitro* [132]. Of note, both RARE and GABP binding motif are highly conserved across the entire PSG family, suggesting members of the

retinoid receptors and Ets transcription factors might potentially contribute to the transcription of other PSG members [132,138].

Besides, the SP1/KLF6-directed *PSG5* activation in trophoblast JEG-3 cells has been reported to be enhanced through inhibiting histone deacetylases (HDACs) [139], providing clues into the potential role of epigenetic factors in PSG expression. Consistently, three consensus binding motifs of the CCCTC-binding factor (CTCF) have been recently identified in the 1.6-1.8 kb regions upstream of all human *PSG* genes, suggesting the CTCF-mediated chromatin looping might also be involved in PSG regulation [140].

On the other hand, the dysregulation of PSGs was found in non-trophoblastic cells. For instance, PSG9 has been recently reported to be upregulated in the tumor tissue and plasma sample of breast cancer patients, in which the canonical TGF-β1/Smad signalling might induce PSG9 activation through the interaction of Smad3 and Smad4 with the *PSG9* promoter [141]. Altogether, responding to a variable external stimulus, PSGs seem to be expressed in a condition-dependent manner. However, whether and how a given PSG member is specially expressed still remains unclear. In addition, the post-transcriptional or other controlling mechanism of human PSGs has been virtually unexplored.

1.2.3 Function of human PSGs

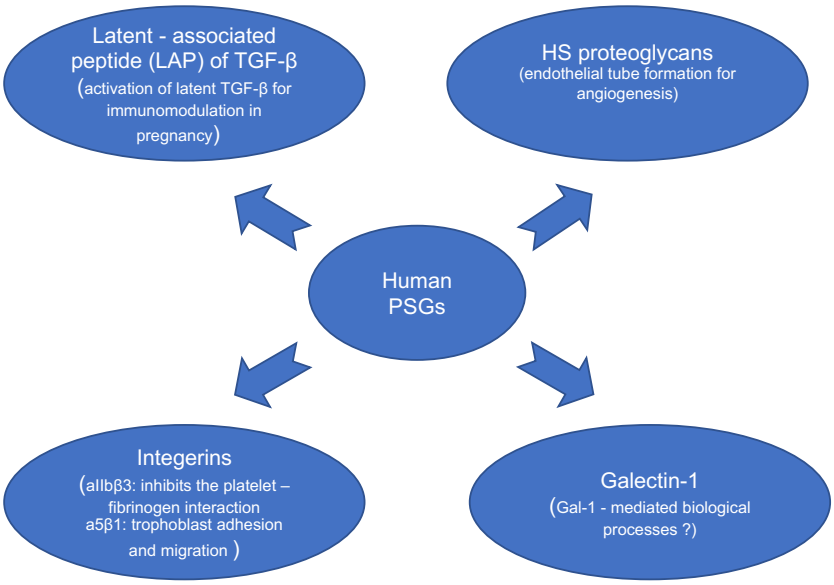

Fig. 5 Diagram depicting the known functions of PSGs in pregnancy.

As a polypeptide hormone abundant in the maternal circulation, PSG proteins seem to function in a context-dependent manner, where PSGs interact with divergent ligands to fine-tune their biological roles in pregnancy (Fig. 5). A notable example of this is the activation of latent TGF-β1 by PSG1 and PSG9, which might be a conserved function across the entire human PSG family [4]. PSGs could activate latent TGF-β1 deposited on the extracellular matrix (ECM) through binding to the latent-associated peptide (LAP) [4,142]. Importantly, the activation of latent TGF-β1 by PSGs could exert biological outcomes in *vivo.* For instance, administration of PSG1 could protect mice from dextran sodium sulfate (DSS)-induced colitis, decrease the production of pro-inflammatory cytokines and increase numbers of T regulatory cells [142]. By contrast, the PSG1-mediated protection is compromised by the addition of neutralizing anti-TGF-β1 antibody [142]. Consistently, TGF-β1 has been well recognized as a cytokine essential for multiple processes in pregnancy, including decidualization, trophoblast invasion, angiogenesis and immune immunomodulation [143-146]. Thus, it has been proposed that PSGs could potentially increase the availability of biologically active TGF-β1 to facilitate establishment of the maternal immunotolerance to the semi-allogeneic fetus during pregnancy [4].

Besides, it has been reported that the binding of the B2 domain of PSG1 to the heparan sulfate proteoglycan (HSPG) on cell membrane is required for endothelial tube formation in *vitro*, suggesting a pro-angiogenic role during human placentation [3,147]. Furthermore, the presence of the tripeptide motif Arg-Gly-Asp (RGD) in the N-domain of most human PSGs led to the hypothesis that similar to the disintegrins in snake venom, PSGs might interact with integrins, disrupting cell-extracellular matrix interactions or modifying other integrin-mediated functions [1,148]. Indeed, human PSG1 has been shown to bind the integrin aIIbβ3 and inhibit the platelet-fibrinogen interaction, likely preventing platelet aggregation and thrombosis in the haemochorial placentation [148]. More recently, PSG transcripts and proteins have been detected in EVTs, despite the predominant expression in SCTs [5,78]. Of note, PSG1 directly interacts with the integrin a5β1 and enhances the adhesion and migration of EVTs in *vitro*, providing new insights into the biological role of PSGs in placentation [5]. On the other hand, as described above, PSG1 has also been also identified to bind Gal-1 in a carbohydrate-dependent manner to protect Gal-1 from oxidative inactivation [126]. Nevertheless, whether this interaction can cause a biological consequence is to be evaluated. Gal-1 has been implicated in several biological processes in pregnancy

including angiogenesis, immunomodulation, cell adhesion, trophoblast syncytialization and invasion [149-151]. Next, it could be of interest to interrogate whether PSG1 has a role in affecting the Gal-1-mediated functions (Fig. 5).

Despite the conserved functions mentioned above, accumulating evidence has shown different PSG members might play different roles in a given context. For instance, in a recent three-dimensional (3D) trophoblast motility assay, the recombinant PSG9 could reduce trophoblast motility whereas PSG1 seems to exert a contrary effect [152]. Furthermore, with a selected reaction monitoring (SRM) mass spectrometry, the increased levels of PSG9 have been detected in maternal plasma of pre-eclamptic women at as early as 15 weeks' gestation, suggesting a potential usage to predict EO- PE [6]. By contrast, PSG1 concentration is decreased in the serum of African American women diagnosed with EO- and LO- PE but only when carrying a male fetus [5]. Altogether, these data suggest the function of different PSG member in a given condition could be variable and needs to be experimentally evaluated case by case.

1.3 Endogenous retroviruses (ERVs) in placenta

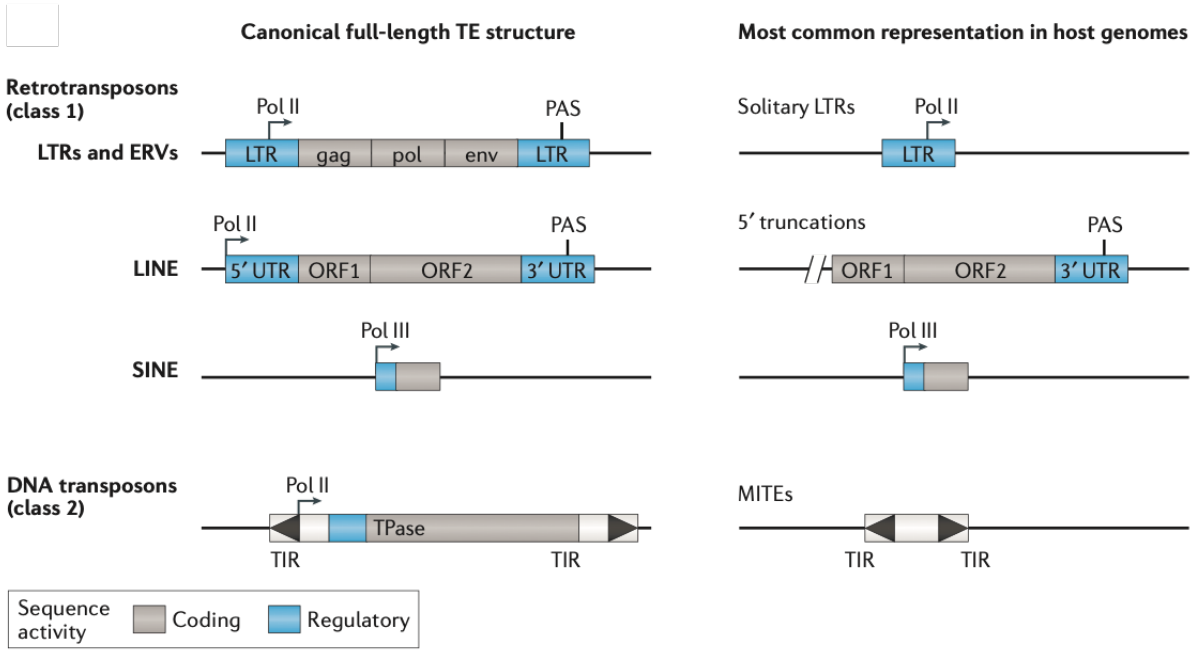

Fig. 6 Diagram depicting the major transposable element classes. The left panel depicts the canonical full-length structure of each TE type. Most TEs contain regulatory sequences highlighted in blue, including the promoters bound to RNA polymerase II or III, and polyadenylation signal (PAS) to control gene expression for transposition. The right panel depicts the commonly displayed organization of each transposon type in host genome. LTRs, long terminal repeats; ERVs, endogenous retroviruses;

LINE, long interspersed nuclear element; SINE, short interspersed nuclear element; TIR, terminal inverted repeat; MITEs, miniature inverted-repeat transposable elements. Adapted from [153] with the license provided by Springer Nature and Copyright Clearance Center.

Nearly all eukaryotic genomes contain a substantial fraction of 'junk DNA': a 'fossil record' of waves of invasions of transposable elements (TEs) [154,155]. TEs are the DNA sequences capable of moving to new loci within the genome (Fig. 6 left panel). With the accumulating mutations or truncations over millions of years, a majority of TEs lost the ability to propagate in modern host genomes (Fig. 6, right panel). However, some of the so-called 'junk DNA' have been appreciated to be transcriptionally active and to act as the 'controlling elements' essential for the host development and health [156-158].

TEs can be generally divided into two major classes based on the transposition intermediates: retrotransposons (class I) and DNA transposons (class II) [159] (Fig. 6). The class I retrotransposons propagate via an RNA intermediate which is then reversely transcribed into a DNA copy and further integrated into the genome. Since the original copy remains in situ, retrotransposons are also known as the 'copy-and-paste' elements [154]. By contrast, the class II DNA transposons mainly mobilize through the 'cut-and-paste' mechanism by which the element excises itself and reinserts in a new genomic location [155].

Retrotransposons can be further subdivided up to the presence or not of the long terminal repeats (LTRs) flanking the internal coding region (Fig. 6). The non-LTR retrotransposons, such as the long interspersed nuclear element -1 (LINE1, L1) and the short interspersed nuclear element (SINE), constitute up to 40% of human genome [155,160,161]. The full-length L1 is an autonomous element that can encode the fully enzymatic machinery necessary for its own transposition. Nowadays, L1 elements make up around 17% of our modern genome with ~ 500,000 mostly fragmented copies, 100 or so of which have been thought still active[155,162]. In contrast, SINEs such as the *Alu* sequences have been recognized as the most prevalent elements in human genome with over 1,000,000 copies [163]. SINEs are nonautonomous elements but still capable of movement in *trans* by 'piggybacking' on the transposition machinery produced by the autonomous transposons [164]. On the other hand, both the full-length LTR retrotransposons and endogenous retroviruses (ERVs) are the autonomous elements with a different replicative strategy from L1 [154,165]. The direct repeat at each end of LTR elements harbors the regulatory sequences essential for expression of the

replicative machinery [166] (Fig. 6). The LTR retrotransposons and ERVs display strikingly similar features and could share the common evolutionary ancestry [167,168]. The only substantive distinction lies in the acquisition of fusogenic *env* genes in ERVs [167]. However, unlike the exogenous retroviruses, a vast majority of ERVs can no longer propagate or encode infectious viruses due to the frequent loss of *env* genes and other mutations accumulated over millions of years [169]. Despite the extinction, some remnants of ERVs can be active and repurposed to a wide range of biological functions in the host [170,171]. Of particular interest, the ERVs-derived activity seems recurrent in placenta [8]. Indeed, in line with the emerging roles of other TEs in host gene expression, mounting evidence supports that ERVs serve as a substantial source of both coding and regulatory sequences in the placental development and evolution [7,172-175] (Fig. 7).

1.3.1 The co-option of ERVs in placenta

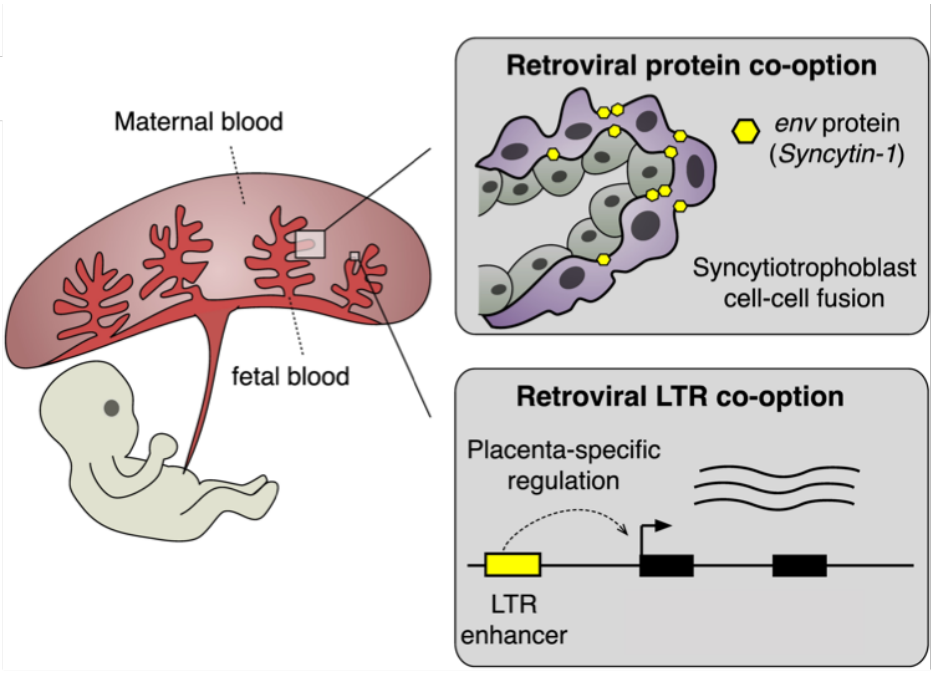

Fig. 7 Diagram depicting the co-option of ERV in placenta. The up panel shows ERV-derived proteins, such as Syncytin-1 for SCT formation. The down panel shows ERV-derived regulatory sequences, such as LTR enhancer for placental gene expression. Modified from [8].

It has been estimated that retroviruses have been infecting vertebrates for over 450 million years [176]. Upon entry into a target cell, the RNA-based viral genome is reverse transcribed into a double-stranded DNA copy which is further integrated into the genome of the host cell [168]. The integrated viral genome, termed provirus, contains two LTRs flanking the canonical retroviral genes, including *gag, pol* and *env*. The LTRs

are enriched with the *cis*-regulatory elements that control proviral gene expression, such as enhancers, promoters and polyadenylation signals, while the retroviral genes encode the structural proteins essential for the replication cycle [168]. Retroviruses typically infect the host somatic cells, whereas the occasional integrations into the germline cells or their precursors could be inherited and further fixed in the host gene pool as the ERVs [168]. Over thousands to millions of years, the ERV-derived sequences have been ubiquitously endogenized in vertebrate genomes, constituting over 8% of the human genome (10% in mice) [161]. ERVs have preserved the original features of their ancestral proviruses to a variable extent, ranging from the complete set consisting of the LTRs and retroviral genes to the highly fragmented remnants of proviruses [168,177]. For instance, around 90% of ERVs in human genome have been thought to exist as the solitary LTRs [161], termed 'solo' LTRs, which could be generated by the homologous recombination between the 5' and 3' LTRs with excision of the intervening sequences [178]. Due to the accumulating sequence mutations as well as the defensive strategies evolved in the host, the vast majority of ERVs have been thought inactive or non-functional. Occasionally, some insertions might confer benefits to the host and have been co-opted (also known as domesticated or exapted) for a biological function as other cellular genes [8,153].

Notably, the placenta has long been recognized to be a 'hotbed' of ERV activity, but the reasons for this have been poorly understood [8]. One explanation is that the relatively permissive chromatin environment for specific ERVs may have co-evolved with placental development [179,180]. For instance, from mouse embryonic day 4.0 to 6.5, when the early cell fate commitment occurs, mouse trophoblast lineage remains globally hypomethylated compared to the epiblast [9,181]. DNA methylation is well known to play critical roles in silencing TEs. Consistently, the promoter activity of several LTRs in placenta has been associated with the lower DNA methylation at the integration sites [182,183]. Despite the inspiring clues, to what extent the recurrent co-option of ERVs in placenta is mediated by DNA hypomethylation is to be systematically evaluated. In addition, it has been reported that unlike primary embryonic stem (ESCs) and extraembryonic endoderm stem (XENs), mouse trophoblast stem cells (TSCs) fail to silence exogenous retrovirus probably due to the absence of specific epigenetic modifiers [180]. It seems understandable that ESCs repress retroviral activity to prevent the potentially heritable detrimental mutations, whereas placental cells could be more

tolerant perhaps because of their transient nature [8,180]. Furthermore, the regulatory activity of LTRs in placenta might also be endowed with their sequence composition which has been repeatedly mutated over the course of evolution [183]. For example, in mouse trophoblast stem cells, the elements from a single ERV family have been evaluated to constitute 35% of the predicted enhancers by providing the binding sites of key trophoblast transcription factors [184]. Whatever the reason, the recurrent exaptation of ERVs in placenta is one of the most suggestive examples showing that TEs could drive their 'selfish' replication by occasionally providing the host with a selective advantage.

1.3.2 ERV-derived proteins in placenta

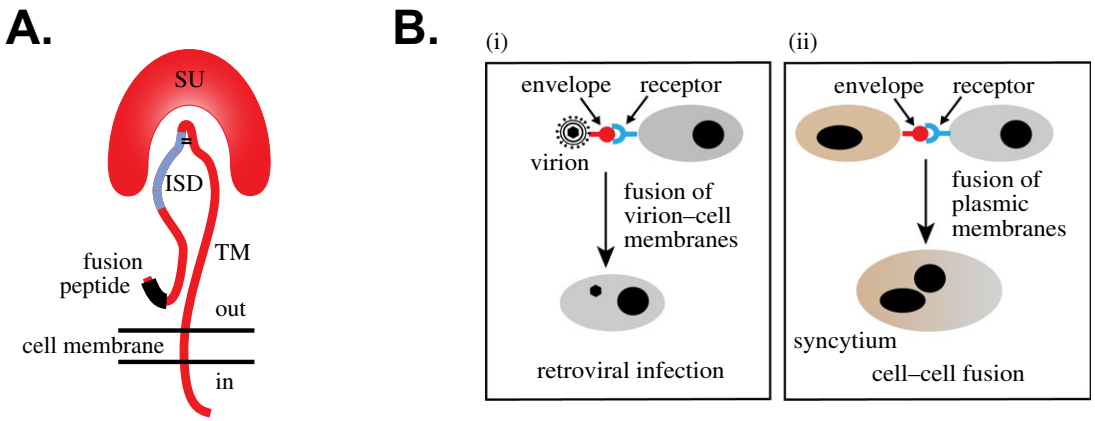

Fig. 8 The retroviral envelop glycoprotein (Env). (**A**) Structural organization of Env at cell surface. (**B**) Diagram depicting that interaction of Env with its receptor can mediate virion-host cell membrane fusion (i) or cell-cell membrane fusion (ii). SU, surface unit; TM, transmembrane unit; ISD, immunosuppressive domain. Adapted from [7] with the license provided by The Royal Society (U.K.).

A large body of evidence has suggested the retroviral *env* gene are repeatedly co-opted to fulfil critical biological functions in the placenta [7]. The *env* gene of retrovirus encodes the envelope glycoprotein (Env) responsible for viral entry into susceptible cells by mediating fusion of the virion envelope with the cell plasma membrane [185]. Interestingly, some proteins expressed by the *env* genes of ERVs in host cells seem to have largely retained the features of their ancestral counterparts, including recognition of a specific cell receptor, fusogenic and immunosuppressive functions (Fig. 8).

One striking example of this is the Syncytin protein. In human placenta, Syncytin-1 and Syncytin-2 are encoded by the *env* genes of HERVW-1 [40,41] and HERVFRD-1

[186], respectively. Structurally, similar to exogenous retroviral Env, the mature Syncytin protein consists of two subunits, surface (SU) and the transmembrane (TM), and displays as homotrimers at the cell plasma membrane [187,188] (Fig. 8A). The SU subunit possesses the receptor-binding domain responsible for recognition and binding to the cell surface receptor, whereas the TM subunit is involved in the cell-cell fusion and immunosuppressive function with three domains: the fusion peptide (FP), the immunosuppressive domain (ISD) and the transmembrane domain (TMD) (Fig. 8A). In human first trimester villi, Syncytin-1 seems to be broadly expressed across different trophoblast subtypes [189] and promotes cell fusion with binding to its cell surface receptors, ASCT-1 and -2 [41,190]. However, Syncytin-2 expression is restricted to a small fraction of VCTs [191], and its receptor, Major Facilitator Superfamily Domain Containing 2 (MFSD2), is localized to the cell membrane of SCTs [42]. Such coordinated expression of Syncytin-2 and MFSD2 in adjacent cells has been proposed to facilitate the polarized 'in-fusion' of SCTs with the underlying VCTs [7,42]. Importantly, the knockdown of either Syncytin-1 or Syncytin-2 leads to a decrease of trophoblast fusion in *vitro* to variable degree [192,193], suggesting an essential role of Syncytin in SCT differentiation. Thanks to the identification of Syncytin-A and -B in mouse [194], the fusogenic capacity of Syncytin to promote SCT formation has been further confirmed in *vivo*. The deletion of Syncytin-A causes defective cell fusion within the syncytiotrophoblast layer I (ST-I) with embryonic lethality at mid-gestation [195], whereas the Syncytin-B null mice display impaired formation of the ST-II layer with relatively viable embryos [194]. Furthermore, the double-knockout leads to more premature death of Syncytin-A null embryos [194], suggesting both Syncytin-A and -B contribute independently to the formation of the two SCT layers during mouse placentation [7,194].

The advantage of the co-opted Env in placenta seems not to be restricted to the intrinsic fusogenic activity. Indeed, it has been proposed the Syncytin-1 and -2 displayed at the surface of placental exosomes might inhibit the production of the Th1 cytokines via the ISD domain [196,197], suggesting a potentially immunomodulatory role in pregnancy. Furthermore, the truncated Env protein lacking the TM unit, termed suppressyn, seems to specially inhibit the Syncytin-1-mediated trophoblast fusion in *vitro* through binding the ASCT-2 receptor [198]. Given that ASCT-2 has been identified as the receptor for a wide family of retroviruses, it could be intriguing to test whether suppressyn has a role in blocking the viral entry into placenta by interfering with the

ASCT-2 binding [198]. In line with the biological functions, the decreased levels of Syncytin have been implicated in human pregnancy disorders such as PE [199], although the underlying etiology is to be clarified. By contrast, staying with its ancestral property which mediates retroviral infection, Syncytin-1 has been recently reported to facilitate HIV infection in trophoblasts by inducing the cell-cell fusion with the HIV positive immune cells [200], highlighting the 'double-edged sword' role of ERV co-option in placenta [153].

Besides the members described here, around a dozen of the *env*-encoded 'Syncytin-like' proteins have currently been identified in placental mammals [201,202]. Interestingly, although they display the similar features such as fusogenic activity, placenta-specific expression and conservation over evolution, these Syncytin members are not clearly orthologous, suggesting they might have been independently captured across different mammals through convergent evolution [7]. Placenta is well known as one of the most variable organs among mammalian species with the difference in degree of trophoblast syncytialization and invasion [203,204]. The stochastic capture of retroviral *env* genes among mammals has been proposed to play a role in shaping the placental morphological diversity [7]. Remarkably, with all canonical characteristics of mammalian Syncytin, the 'Syncytin-Mab1' has been recently identified in the nonmammalian *Mabuya* species which possesses the placental structure, indicating the Syncytin capture is not restricted to placental mammals and could potentially be a driving force for placental emergence and evolution [205].

1.3.3 ERV-derived regulatory elements in placenta

Besides the co-option as proteins, TEs-derived sequences, particularly the LTRs of ERVs, have been recently recognized to be a rich source of *cis*-regulatory elements for gene expression, providing promoters, enhancers, insulators, splicing and polyadenylation sites [153,206]. For instance, with the relatively low levels of DNA methylation, an LTR of the THE1D retroelement family could drive the placenta-specific expression of the interleukin-2 (IL-2) receptor β subunit (IL-2Rβ) as an alternative promoter [182]. More recently, the regulatory role of LTR-derived elements as species- and tissue-specific enhancers has been well characterized. One pioneering study done by Chuong *et al.* has revealed the RLTR13-derived elements function as mouse-specific enhancers in placenta, highlighting the role of ERVs in

rewiring the placental gene regulatory network [184]. Strikingly, of the 608 RLTR13D5 copies, 95 exhibit the enhancer potential marked by the monomethylation of histone H3 at lysine 4 (H3K4me1) and acetylation of histone H3 at lysine 27 (H3K27ac) in mouse TSCs [184,207]. In addition, the RLTR13D5 copies with enhancer potential bind to the core transcription factors defining mouse TSCs, including Eomes, Cdx2 and Elf5 [208-210], indicating the LTRs could be exapted as mouse-specific enhancers by coordinating the core TSC regulatory network for placental development [184]. Consistently, the predicted RLTR13D5-derived enhancers could functionally drive gene expression in a luciferase reporter assay [184]. Despite the inspiring results, one crucial question unexplored is whether the LTR-derived enhancers are essential for endogenous expression of the target genes, because the biochemical hallmark at the chromatin level seems not be always to be indicative of regulatory function. Indeed, a comprehensive survey performed the CRISPR interference of RLTR13D6 elements in mouse embryonic stem cells and revealed that only a minority of copies might have significant roles in gene regulation [211], emphasizing that the regulatory function of LTR-derived elements should be experimentally examined in their native chromatin context.

Functional evaluation of a given TE in situ has been hampered partially due to their repetitive property. The recent advance of precise gene-editing technologies such as CRISPR/Cas9-based methods has provided powerful tools to address this. For example, the CRISPR/Cas9-mediated deletion of *Enhancer L*, a LTR71-derived distal enhancer, results in complete ablation of HLA-G expression in human trophoblasts, suggesting *Enhancer L* might be a bona fide enhancer driving the placenta-specific expression of HLA-G [212]. Furthermore, the essential role of LTR-derived enhancer in the placental expression of corticotropin-releasing hormone (CRH) in primates has been also reported. Introduction of the human CRH cluster containing a LTR enhancer into mice could induce ectopic expression of CRH in placenta, leading to the prolonged gestational length [213]. Importantly, removal of the LTR enhancer by CRISPR method could downregulate CRH expression in mice placenta and fully rescue the gestation length back to control levels [8,213]. Altogether, these data support the notion that the co-option of ERVs could contribute to shaping the host gene regulatory networks and further, at least in some contexts, exert a biological consequence.

Despite the exciting advances, the evolutionary mechanism and significance underlaying the frequent co-option of ERVs in placenta have been far from being well understood. For instance, the vast majority of ERV-derived elements in placenta are

to be functionally assessed [8]. In particular, the regulatory role of ERVs in gene expression at post-transcriptional level has rarely been explored [214,215]. On the other hand, given mounting cases showing the pathogenic effects of TEs in human disease, it is tempting to interrogate whether the ERV-derived regulatory activity plays a role in pregnancy complications.

1.4 Project objectives

PSGs have been recognized to be mainly expressed in placenta and secreted into maternal circulation for multiple physiological processes during pregnancy, including angiogenesis, immunomodulation and trophoblast migration. The abnormal levels of PSGs in maternal blood have been detected in PE. Despite the correlations, either the regulation or biological role of PSG expression in trophoblast differentiation has been poorly understood. Mounting evidence has suggested that the ERV-derived proteins or regulatory sequences are frequently co-opted in placenta. The first objective of this project is to investigate the regulatory role of human ERVs (HERVs) in the expression of PSG9, a proposed predictive biomarker of EO-PE. To this end, we screened the candidate HERV elements which potentially possess regulatory activity through mining the online available epigenetic data as well as the experimental methods. Next, given the robust expression of PSGs in SCTs, we interrogated the biological function of HERVs-directed PSG9 regulation in SCT formation. Lastly, we examined the expression of PSGs in the trophoblasts of EO-PE and explored the pathological role of PSG9 misexpression in this disease. Altogether, our project identified two HERVs-derived elements, LTR8 and MER65-int, which could modulate PSG9 expression at transcriptional and post-transcriptional levels, respectively. Importantly, the HERVs-directed expression of PSG9 could play an essential role in placentation through facilitating trophoblast syncytialization. Furthermore, the HERVs-directed regulation contributes to the generation of the PSG9 isoform. Of note, the secretory PSG9 was identified to be upregulated in trophoblasts of EO-PE, supporting elevated PSG9 as a potential biomarker of this disease.

2. Materials and Methods

2.1 Plasmid construction

a. The corresponding coding sequences were amplified and introduced into the vectors digested with appropriate restriction enzymes by the NEBuilder HiFi DNA Assembly Master Mix (#2621 NEB) (Table.1).

Table. 1 Reaction mixture for PCR fragments assembly (20 μl reaction system)

Fragments	Vector digested with restriction enzymes	Candidate fragment
Volume (μl)	x (0.06 pmols)	y (0.12 pmols)
2x NEBuilder HiFi DNA Assembly Master Mix (μl)	10	
Deionized H2O (μl)	10 - x - y	
Incubation	1 hr at 50 °C	

b. 2 μl of the chilled assembled product was used for bacterial transformation. Then the resulting plasmids were confirmed by sequencing and purified with the NucleoBond Xtra Midi kit (#740410.50 MACHEREY-NAGEL).

2.2 Cell culture

BeWo trophoblast cell line was developed from a malignant gestational choriocarcinoma of the human placenta [216]. BeWo cells were maintained in GlutaMAX™ DMEM/F-12 (#31331028 Thermo Fisher Scientific) medium supplemented with 10% (v/v) Foetal Bovine Serum (FBS) and 1% (v/v) penicillin/streptomycin (P/S). The Saint Georges Hospital Placental Line-4 (SGHPL4) is a cell line of the human extravillous trophoblast in the first trimester [217]. SGHPL4 cells were cultured in the Ham's F-10 liquid medium (#FG0715) supplemented with 10% FBS and 1% P/S. Both BeWo and SGHPL4 cell lines were routinely cultured in 5% CO_2 at 37°C.

For routine subculture, the cells at ~ 90% confluency were split with the 0.25% (w/v) Trypsin-0.53 mM EDTA solution and harvested. An appropriated aliquot of cell suspension was transferred to new culture vessels.

For cryopreservation, cell suspension was aliquoted with the serum-freezing solution into cryovials (1 x 10^6 cells per vial at least) and kept in a freezer box at -80°C

for 24 hours (hr). Then the cryovials were transferred to liquid nitrogen for long-term storage.

2.3 Cell transfection

2.3.1 BeWo cell transfection

BeWo cell transfection was done with Lipofectamine™ 3000 Reagent (#L3000015 Thermo Fisher Scientific) following the manufacturer's instructions.

a. 1st day, 1x10^6 BeWo cells were seeded and cultured for 24 hr in each well of 6-well plate before transfection.

b. 2nd day, BeWo cells were transfected according to the manufacturer's instructions (Table. 2).

c. 4th day, Transfected cells were incubated for 48 hr. (Optional: To obtain the stable cell lines with the *Sleeping Beauty* (*SB*) system [218], transfected cells were further selected with the complete medium containing ~ 4 μg/ml Puromycin for 10 ~ 14 days.)

Table. 2 Reagent mixture for BeWo cell transfection. Each reaction mix volume was prepared for one well of 6-well plate and accounted for pipetting loss. Volumes were scaled proportionally for additional wells.

	Component	Volume/Amount
	BeWo	1x10^6
Step 1	Opti-MEM ™ Medium	125 µl
	Lipofectamine ™ 3000 Reagent	3.75 µl
Step 2	Opti-MEM ™ Medium	125 µl
	Plasmids	~2,5 µg
	P3000 ™ Reagent (2 µl/µg DNA)	5 µl
Step 3	Add Diluted DNA to each tube of Diluted Lipofectamine ™ 3000 Reagent (1:1 ratio)	
	Diluted Lipofectamine ™ 3000 Reagent	125 µl
	Diluted DNA (with P3000 ™ Reagent)	125 µl
Step 4	Incubate for 15 min at room temperature, then add DNA-lipid complex to cells	

2.3.2 SGHPL4 cell transfection

SGHPL4 cell transfection was performed with the Neon™ Transfection System (#MPK5000 Thermo Fisher Scientific) following the manufacturer's instructions.

a. The ~ 90% confluent cells were washed twice with DPBS, treated with the 0.25% (weight/volume or w/v) Trypsin-0.53 mM EDTA solution and collected in the complete medium. 10 µl cell suspension was pipetted for cell counting.
b. Cells were centrifuged at 300 x g for 5 min and cell pellet was resuspended in the resuspension Buffer R (1~3 x 10^6 SGHPL4 cells per 100 µl Buffer R).
c. Cell suspension was electroporated with plasmids using the parameters: 1260V/20ms/2 as pulse voltage/pulse width/pulse number, respectively.
d. Electroporated cells were cultured in the 10 cm dish containing the medium without antibiotics for 24 hr. (Optional: To obtain the stable cell lines with the *SB* system, transfected cells were further selected with the complete medium containing ~ 4 µg/ml Puromycin for 10 ~ 14 days.)

2.4 Overexpression of gene of interest (GOI)

2.4.1 Transient overexpression of GOI in BeWo cells

a. Following the protocol described in section 2.1, the GOI was introduced into the *SB* vector (5'-IR-CAGGS^{prom}-HA-MCS-SV40^{prom}-*Puro*-3'IR. MCS, multiple cloning site; IR, inverted repeats) which was digested with EcoR I and Not I restriction enzymes.
b. Cells were seeded onto 6-well plates (1 × 10^6 cells/well) for 24 hr.
c. Cells were transfected with plasmids (~ 2.5 µg per well) by the Lipofectamine™ 3000 Reagent following the protocol mentioned above.
d. At 24 hr post-transfection, transfected cells were selected with the completed medium containing ~ 4 µg/ml Puromycin for 48 hr.
e. Transfected cells were harvested. Then gene overexpression was confirmed by RT-qPCR or western blot. The cells expressing the mCherry fluorescent protein which was produced in the same manner were used as a negative control.

2.4.2 Stable overexpression of GOI in BeWo cells

To construct the cell lines stably expressing GOI, the experimental protocol was as same as that for transient overexpression except that cells were co-transfected with the expression cassette of GOI (2.5 µg per well) and the plasmid pcGlobin2- *SB100X* (250 ng per well) which expresses the hyperactive SB transposase.

2.5 Quantitative reverse transcription PCR (RT-qPCR)

2.5.1 RT-qPCR with human trophoblast cell lines

a. The total RNA extraction was carried out with the Direct-zol™ RNA MiniPrep Kit (#R2025 Zymo Research) following the manufacturer's instructions.
b. 500 ng ~ 1 µg total RNA was used to synthesize the single-stranded cDNA with the High-Capacity RNA-to-cDNA™ Kit (#4387406 Thermo Fisher Scientific) following the manufacturer's instructions.
c. The resulting cDNA was stored at -20°C or used for real-time qPCR with the SsoAdvanced™ universal SYBR® Green Supermix (#1725271 Bio-Rad) on the CFX96 Touch™ system (Bio-Rad) (Table 2 and 3).
d. Gene expression was analysed using the CFX Maestro™ Software (Bio-Rad) with the ΔΔCT method and normalized to human *ACTIN*.

Table. 3 Reaction setup for RT-qPCR. (10 µl reaction system, volumes were scaled proportionally for additional reactions.)

Component	Volume per 10 µl reaction	Final Concentration
SsoAdvanced™ universal SYBR® Green Supermix (2 x)	5 µl	1 x
Forward and reverse primer mix (10 x)	1 µl	500 nM each
cDNA template	0.1 µl	
Nuclease-free H_2O	3.9 µl	

Table. 4 Thermal cycling protocol for real-time qPCR

Polymerase activation	Denaturation	Annealing/ Extension	Cycles	Melt-curve
95°C, 30 sec	95°C, 15 sec	60°C, 60 sec	39	65–95°C, 0.5°C increment 2–5 sec/step

2.5.2 RT-qPCR with human placenta

Human placenta sampling was approved by the Regional Committee of the Medical Faculty of Charité Berlin. The RT-qPCR was done following the protocol from [219] in collaboration with Dr. Florian Herse from the Müller/Dechend Lab at Experimental and Clinical Research Center (ECRC).

Table. 5 Clinical characteristics of the placenta samples for RT-qPCR analysis. Early is defined as delivery earlier than 34 gestational weeks and late is defined as delivery later than 34 gestational weeks. Data are presented as mean ± standard deviation (mean ± SD). Unpaired t test, cases vs. controls, *$p<0.05$, **$p<0.001$. PE, pre-eclampsia; IUGR, Intrauterine growth restriction; BMI, body mass index.

Characteristics at delivery	Control early (n=7)	PE early (n=7)	IUGR early (n=8)	Control late (n=9)	IUGR late (n=5)	PE late (n=18)
Maternal age (years)	30.7 ± 7.0	28.6 ± 2.2	27.4 ± 6.9	34.9 ± 5.5	29.6 ± 6.5	30.6 ± 5.7
BMI (kg/m^2)	26.1 ± 8.2	26.0 ± 6.1	25.3 ± 5.7	23.0 ± 5.1	26.1 ± 6.1	28.0 ± 6.9
Gestational age (days)	210.6 ± 23.6	204.9 ± 20.5	213.8 ± 23.0	266.3 ± 15.1	272.6 ± 7.8	255.3 ± 12.9
Blood pressure Systolic (mm Hg)	110.0 ± 10.0	153.3 ± 9.9**	122.9 ± 13.8	113.9 ± 12.6	100.0 ± 14.1	153.1 ± 15.3**
Blood pressure Diastolic (mm Hg)	68.6 ± 6.9	96.8 ± 10.2**	80.7 ± 11.7	70.6 ± 9.5	66.0 ± 9.0	95.3 ± 9.8**
Albuminuria (yes/no)	0/7	7/0	2/6	0/9	0/5	18/0

2.6 Western blot

2.6.1 Protein extraction

a. Cell pellet was incubated at 4°C for 2 hr with the lysis buffer containing 50 mM Tris-HCl PH 8.0, 100 mM NaCl, 5% Glycerin, 10 mM EDTA, 1% NP-40 and an appropriated amount of protease inhibitors (#A32955 Thermo Fisher Scientific).

b. Cell lysate was centrifuged at 14,000 rpm for 10 min at 4°C and the supernatant was harvested into a 1.5 ml microcentrifuge tube.

c. Protein concentration was measured with the Pierce™ BCA™ Protein Assay (#23225 Thermo Fisher Scientific).

2.6.2 SDS-PAGE and western blot

a. Protein sample was denatured in the SDS loading buffer (5x stocking solution: 0.25% bromophenol blue, 0.5 M dithiothreitol, 50% glycerol, 10% sodium dodecyl sulfate, 0.25 M Tris-Cl pH 6.8) by heating reaction at 96°C for 10 min and loaded in SDS-PAGE gels (10 wells, ~ 20 µg per well).

b. The SDS-PAGE running was done with the TGX Stain-Free™ FastCast™ acrylamide kit (10% #1610183 Bio-Rad) following the manufacturer's instructions.

c. Protein transfer was performed by the Trans-Blot® Turbo™ Transfer System (Bio-Rad) with the running parameters: 1.3 A for 10 min.

d. The transferred membrane was blocked with the 5% skim milk-TBST buffer (w/v) for 1 hr at room temperature (RT)

e. The membrane was incubated with the 1st antibody for overnight at 4°C.

f. The membrane was washed with the 1 X Tris-buffered saline with 0.1% Tween® 20 detergent (TBST buffer) for 6 times, 5 min each, and incubated with the 2nd antibody for 1 hr at RT.

g. The membrane was washed with the TBST buffer for 6 times, 5 min each. The chemiluminescence signal was developed with the ECL™ detection reagent (#RPN2232 GE Healthcare) and captured under the Bio-Rad ChemiDoc™ MP Imaging System.

2.7 Cell fractionation assay

To examine the cellular localization of PSG9 isoforms, we constructed the BeWo cell lines co-expressing two representative PSG9 isoforms which were fused with distinct tag proteins (see the diagram in the results section). The transfected cells were used for the cell fractionation assay.

2.7.1 Collection of cell culture supernatant

Cells were seeded in 6-well plate (1 X 10^6 per well) and cultured for 48 hr. Then 4 ml culture media were collected and centrifuged at 4000 rpm for 15 min at 4°C. The supernatant was further concentrated using the Amicon® Ultra-15 Centrifugal Filter Unit (#UFC903008, Merck)

2.7.2 Isolation of total membrane proteins.

Cellular membrane and cytoplasmic components were separated by the MinuteTM Plasma Membrane Protein Isolation and Cell Fractionation Kit (#SM-005 Invent Biotechnologies) following the manufacturer's instructions.

a. Filter cartridges were placed in collection tubs and incubated on ice.

b. The ~ 1 X 10^7 cells were collected into the 1.5 ml microcentrifuge tube and centrifuged at 300 X g for 5 min at 4°C. Then cells were washed with cold DPBS once.

c. Supernatant was removed and the cell pellet was resuspended in 500 µl fractionation buffer A. Then cell suspension was incubated on ice for 5 min.

d. The tube was vortexed vigorously for 10 seconds and cell suspension was immediately transferred to the cold prefilter cartridge.

e. The filter cartridge was centrifuged at 14,000 rpm for 30 seconds at 4°C. Cell pellet in the collection tube was resuspended, re-passed the same filter and was centrifuged once more.

f. The filter was discarded and the cell pellet was resuspended by vigorously vortexing for 10 seconds.

g. The collection tube was centrifuged at 3000 rpm (700 X g) for one min (the intact nuclei fraction). Then the supernatant was transferred to a fresh 1.5 ml microcentrifuge tube and centrifuged at 16,000 X g for 30 min at 4°C.

h. The supernatant (the cytoplasmic fraction) was collected into a new tube and the pellet (the total membrane protein fraction, including organelles and plasma membranes) was dissolved into 200 µl detergent-containing buffer for the downstream analysis.

2.7.3 Cell fractionation examination by western blot

Western blot was done following the methods aforementioned except that the transferred membrane was incubated for overnight at 4°C with the 1st antibody mixture containing anti-HA and anti-FLAG. Correspondingly, the membrane was then incubated with a 2nd antibody mixture.

2.8 Immunofluorescence staining

2.8.1 Cell culture and slides preparation

Cells were seeded onto coverslips into 12 well plate (4 X 10^5 per well) for 48 hr.

2.8.2 Fixation

Cells on coverslips were fixed with 4% paraformaldehyde in PBS for 15 min at RT and washed once with cold DPBS.

2.8.3 Permeabilization (optional)

For intracellular staining, cells were permeabilized with 0.2% triton X-100 in PBS for 15 min at RT and washed three times with DPBS.

2.8.4 Blocking and antibody incubation

a. Cells were blocked for 1 hr at RT with 10% serum from the species that the secondary antibody was developed.

b. Cells were incubated for overnight at 4 °C with the 1st antibody. Then coverslips were washed three times with DPBS.

c. Cells were incubated with the 2nd antibody for 1 hr at RT. Then coverslips were washed three times with DPBS.

2.8.5 Nuclear staining

Cells were incubated with the blocking solution containing 1 µg/ml Hoechst 33342 (#H3570 Thermo Fisher Scientific) for 5 min at RT and washed with DPBS once.

2.8.6 Slides mounting

Each coverslip was mounted with ~ 15 µl mounting medium (#H-1000 Vector Laboratories) and was sealed with the sealant (#23005 BIOTIUM) to prevent drying and movement under microscope. Slides were stored in dark at -20°C.

2.8.7 Imaging

Images were captured under the Leica TCS SP8 with a 63x oil/NA 1.4 objective at the Advanced Light Microscopy (ALM) technology platform at MDC.

2.9 Luciferase reporter assay

2.9.1 The enhancer reporter assay

The enhancer activity of LTR8 and other transposable elements at *PSG9* locus (GRCh37/hg19) was measured by a luciferase reporter system.

a. Candidate elements were introduced into the pGL3 Promoter Vector MCS-SV40^{prom}-hluc-SV40pA (#E1761 Promega) which was digested with Nhe I and Xho I restriction enzymes. (Plasmids were constructed by Dr. Julianna Rugor.)

b. For transient transfection, BeWo cells seeded onto 96-well plates (2 X 10^4 cells/well) for 24 hr were co-transfected with the internal control plasmids pRL-SV40 (#E2231 Promega, 50 ng/well) and the candidate plasmids (250 ng/well) using the jetPRIME®transfection reagent (#114-15 Polyplus Transfection) following the manufacturer's instructions.

c. At 48 hr post-transfection, transfected cells were harvested and lysed.

d. The resulting lysate was used to measure the *firefly* (luc) and *Renilla* (Rluc) luminescence with the Dual-Glo® Luciferase Assay System (#E2920 Promega) on the Tecan Spark 10M multimode plate reader according to the manufacturer's instructions.

e. The average luc/Rluc luminescence ratio of at least 6 technical replicates was calculated to represent the relative enhancer activity of candidate elements. The empty vector and the pGL3-Control Vector containing a SV40 enhancer (#E1741 Promega) were used as the negative and positive control, respectively.

2.9.2 The polyadenylation reporter assay

The polyadenylation efficiency of MER65-int and other candidate elements was measured by a recently developed bicistronic reporter system [220].

a. Following the protocol described in the section 2.1, candidate elements were cloned into the dual luciferase vector SV40^{prom}-Rluc-MCS-IRES-luc-SV40pA which was digested with Xho I and EcoR I restriction enzymes. (Plasmid was kindly offered by Prof. Dr. Xianchun Li).

b. For transient transfection, BeWo cells seeded onto 96-well plates (5 X 10^4 cells/well) for 24 hr were transfected with the candidate plasmids (400 ng/well) using the jetPRIME® transfection reagent following the manufacturer's instructions.

c. At 48 hr post-transfection, transfected cells were harvested and lysed.

d. The resulting lysate was used to measure the *firefly* (luc) and *Renilla* (Rluc) luminescence with the Dual-Glo® Luciferase Assay System on the Tecan Spark 10M multimode plate reader according to the manufacturer's instructions.

e. The average luc/Rluc luminescence ratio of at least 7 technical replicates was calculated to represent the relative polyadenylation efficiency of candidate elements. The empty vector and the one containing a known synthetic pA site (SPA) were used as the negative and positive control, respectively.

2.10 Identification of polyadenylation *cis*-element in MER65-int

a. Each MER65-int sequence fused with the 3' terminus of PSG transcript was retrieved as a query sequence (200 bp long, centered at polyadenylation site).

b. The query sequence was used to identify its corresponding alignment in MER65-int consensus sequence.

c. The identification of polyadenylation *cis*-element was performed following the protocol from [221]. The analysis was done in collaboration with Dr. Bin Zhang.

2.11 CRISPR-directed LTR8 deletion in BeWo cells

a. Two pairs of sgRNAs flanking the LTR8 at the human PSG9 locus were designed with the online tool CRISPOR (http://crispor.tefor.net/) [222] and cloned into the pU6-(BbsI) sgRNA_CAG-Cas9-venus-bpA (Addgene plasmid # 86986, a gift from Ralf Kuehn) following the protocol from [223].

b. With the Lipofectamine™ 3000 Reagent, the 5' sgRNA and 3' sgRNA plasmids were co-transfected into BeWo cells in 6-well plates (1.25 µg/well each).

c. At 48 hr post-transfection, the mVenus positive cells were sorted and seeded into 6-well plates at ~ 5000 cells per well.

d. After 10 ~ 14 days cell culture, single colonies were manually picked under microscope and genotyped using the DNA isolated with the QuickExtract™ DNA Extraction Solution (#QE09050 Lucigen).

d. The colonies harboring the homozygous deletion were further confirmed by sequencing and were selected for the downstream analysis. The wide type bulk cells were used as the negative control.

2.12 BeWo cell syncytialization assay

2.12.1 BeWo syncytialization assay (routine version)

a. BeWo cells were seeded in 6-well plates (1 x 10^6 cells/well) for 24 hr.

b. BeWo cells were treated with the culture media containing 50 μM Foskolin (#F6886 Sigma-Aldrich) for 48 hr.

c. Cells were harvested at 24, 48 and 72 hr post-treatment for downstream analysis.

2.12.2 Quantification of BeWo syncytialization

a. The BeWo cell lines stably expressing either the GFP or mCherry were constructed.

b. The BeWo cells expressing GFP were mixed with those expressing mCherry and seeded into 12-well plates (1.5 x 10^5 cells/cell type/well) for 24 hr. The wide type and the BeWo cells expressing either GFP or mCherry alone were seeded (3 x 10^5 cells/cell type/well) as controls.

c.The culture media was replaced with the fresh medium containing 50 μM Foskolin.

d. At 48 hr-post Forskolin treatment, cells were treated with the trypsinization solution for 20 min at 37°C and harvested into the completed media. The cell suspension was gently pipetted 30 times to decrease cell clumping.

e. The cell suspension was centrifuged at 300 X *g* for 5 min and washed with DPBS once.

f. The cell pellet was resuspended into 600 μl FACS buffer (DPBS containing 1% BSA and 5 mM EDTA) and filtered through the 70 μm Flowmi™ Cell Strainers (#15342931 Fisher Scientific) immediately before Fluorescence-Activated Cell Sorting (FACS) analysis.

g. The FACS was performed with the BD LSRFortessa™ cell analyzer on the Flow Cytometry platform at MDC. 20 ~ 30 thousands of single cells were analyzed for each sample and data analysis was done using the FlowJo™ software.

2.12.3 BeWo syncytialization assay with PSG9 KD (modified version)

a. BeWo cells were seeded in 6-well plates (1 X 10^6 cells/well) for 24 hr.

b. BeWo cells were transfected with the plasmid pLKO.1-PSG9shRNA (MISSION® shRNA TRCN0000244632). The pLKO.1-scramble shRNA (#1864 Addgene, a gift from David Sabatini) was used as a negative control.

c. At 24 hr post-transfection, the transfected cells were screened with the culture medium containing 4 μg/ml puromycin.

d. After another 24 hr post-screening, transfected cells were treated with the mixed medium containing 4 μg/mL puromycin and 50 μM Foskolin for syncytialization for 48 hr.

e. Cell fusion was evaluated based on the expression of the signature genes of trophoblast syncytialization.

2.13 Transcriptome of LTR8 KO and PSG9 KD in cell syncytialization

2.13.1 RNA sample preparation

a. The BeWo syncytialization assay with LTR8 knockout (LTR8 KO) was carried out using the routine protocol as described above. The total RNA was extracted from the wild type control and LTR8 KO cells (sgRNA 392 + sgRNA 1073, colony P6-B4) at 48 hr post-Forskolin treatment.

Table. 6 The RNA samples of LTR8 KO for RNA sequencing.

Biological replication	Cell passage No.	RNA prep. Date
#1	Contr. P237 VS. KO. P238	09.03.2020
#3	Contr. P240 VS. KO. P241	19.03.2020
#4	Contr. P241 VS. KO. P242	27.04.2020

b. The BeWo syncytialization assay with PSG9 knockdown (PSG9 KD) was carried out using the modified protocol as described above. The total RNA was extracted from the shRNA scramble control and PSG9 KD cells at 48 hr post-Forskolin treatment.

Table. 7 The RNA samples of PSG9 KD for RNA sequencing.

Biological replication	Cell passage No.	RNA prep. Date
#1	P226	02.06.2020
#3	P228	08.06.2020
#4	P229	15.06.2020

c. The quality control (QC) of all RNA samples, including concentration, 28S/18S ration and RNA integrity number (RIN) were performed with the Agilent 2100 Bioanalyzer System.

Table. 8 Quality control of LTR8 KO and PSG9 KD RNA sample

Sample NO.	Sample Name	Concentration (ng/μl)	Volume (μl)	28S/18S	RIN
S1	Control-1	127	10	1.7	9.3
S2	LTR8KO-1	100	10	1.3	9.4
S5	Control-3	244	10	1.4	9.1
S6	LTR8KO-3	189	10	1.6	9.6
S7	Control-4	111	10	1.8	9.1
S8	LTR8KO-4	57	10	2.4	8.2
S9	Scramble-1	125	10	1.4	8.8
S10	PSG9KD-1	146	10	1.8	9.3
S13	Scramble-3	266	10	1.7	8.5
S14	PSG9KD-3	130	10	1.7	8.6
S15	Scramble-4	69	10	1.6	7.6
S16	PSG9 KD-4	72	10	1.7	8

2.13.2 Sequencing pipeline

The transcriptome of 12 RNA samples were prepared with the stranded mRNA library and sequenced on the ‘DNBseq’ platform at BGI Group (Fig. 9).

a. mRNA enrichment and purification with Oligo dT selection

b. RNA fragment and reverse transcription. For stranded specific mRNA library preparation, the second-stranded cDNA was synthesized with dUTP.

c. End repair, A addition and adaptor ligation

d. PCR amplification

e. Single strand separation and cyclization

f. DNA nanoball synthesis

g. Sequencing on ‘DNBseq’ platform (PE100: paired-end reads, 100bp per read)

h. Data production. After sequencing, the raw reads were filtered, including removing adaptor sequences, contamination and low-quality reads. Around 30M reads per sample (6G clean data) were finally obtained and saved as FASTQ files.

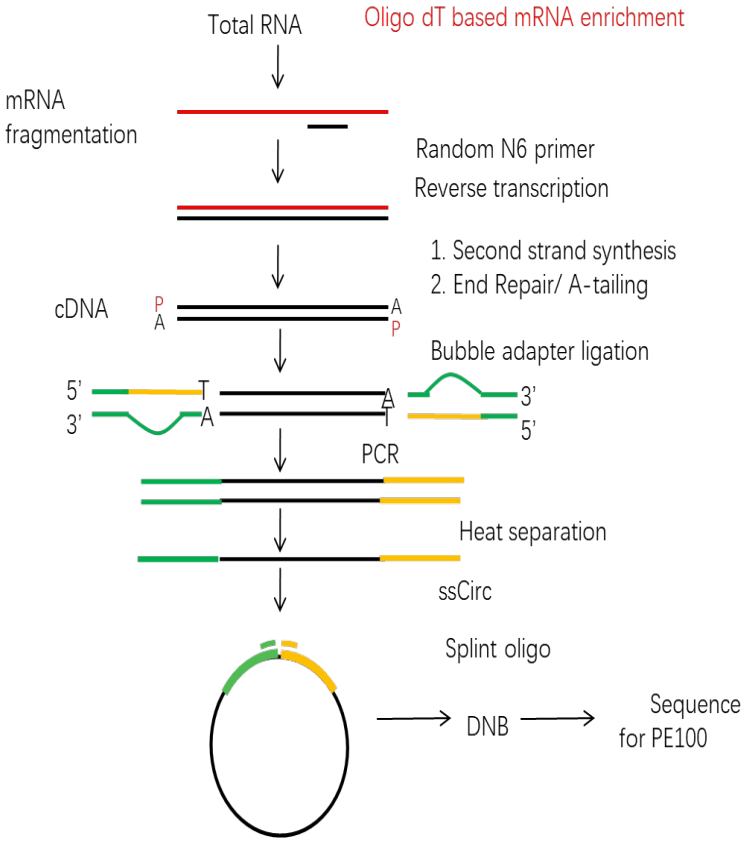

Fig. 9 Sequencing pipeline. Adapted from the "Experimental pipeline of transcriptome" at BGI.

2.13.3 Data analysis

a. Quality control of raw reads

The reads quality was examined using 'FastQC' and 'MultiQC' [224] with the scripts following http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

```
fastqc path/to/sample.fq.gz
multiqc path/to/folder/*_fastqc.zip
```

Table. 9 An overview of general statistics of sequence reads

Sample	% Duplicate Reads	% GC	Length	% Failed in FastQC	Reads (millions)
S1_Control-1_1	68.9%	50%	100 bp	27%	37.3
S1_Control-1_2	69.8%	51%	100 bp	27%	37.3
S2_LTR8KO-1_1	67.8%	49%	100 bp	27%	37.3
S2_LTR8KO-1_2	67.9%	50%	100 bp	27%	37.3
S5_Control-3_1	72.4%	50%	100 bp	27%	36.1
S5_Control-3_2	72.1%	51%	100 bp	27%	36.1
S6_LTR8KO-3_1	69.6%	49%	100 bp	27%	37.3
S6_LTR8KO-3_2	69.5%	50%	100 bp	27%	37.3
S7_Control-4_1	68.6%	50%	100 bp	27%	37.3
S7_Control-4_2	68.4%	51%	100 bp	27%	37.3
S8_LTR8KO-4_1	67.8%	50%	100 bp	27%	37.3
S8_LTR8KO-4_2	67.4%	51%	100 bp	27%	37.3
S9_Scramble-1_1	67.3%	50%	100 bp	27%	32.2
S9_Scramble-1_2	68.0%	51%	100 bp	27%	32.2
S10_PSG9KD-1_1	68.5%	50%	100 bp	27%	37.2
S10_PSG9KD-1_2	68.5%	51%	100 bp	27%	37.2
S13_Scramble-3_1	71.3%	50%	100 bp	27%	37.2
S13_Scramble-3_2	71.5%	51%	100 bp	27%	37.2
S14_PSG9KD-3_1	73.7%	50%	100 bp	27%	37.2
S14_PSG9KD-3_2	74.0%	51%	100 bp	27%	37.2
S15_Scramble-4_1	72.7%	50%	100 bp	27%	37.2
S15_Scramble-4_2	72.7%	51%	100 bp	27%	37.2
S16_PSG9KD-4_1	70.2%	50%	100 bp	27%	37.3
S16_PSG9KD-4_2	71.0%	51%	100 bp	27%	37.3

Fig. 10 Mean quality value across each base position in the read.

b. Reads alignment to reference human genome

Reads were aligned to the GRCh37/hg19 reference genome using the aligner 'Spliced Transcripts Alignment to a Reference (STAR)' [225]. The alignment quality was further assessed by 'RSeQC' package [226].

- Generating genome indexes files

```
--runThreadN 4 \
--runMode genomeGenerate \
--genomeDir /path/to/genomeDir \
--genomeFastaFiles /path/to/ GRCh37.p13.genome.fa \
--sjdbGTFfile /path/to/annotations.gff3 \
--sjdbGTFtagExonParentTranscript Parent \
--sjdbOverhang 99
```

- Mapping paired-end reads (e.g. S1_Control-1) to the hg19 reference genome

```
--runThreadN 4 \
--genomeDir /path/to/genomeDir \
--readFilesIn /path/to/S1_Control-1_1.fq.gz /path/to/S1_Control-1_2.fq.gz \
--readFilesCommand zcat \
--outSAMtype BAM SortedByCoordinate \
--outFileNamePrefix /path/to/output folder
```

Table. 10 Reads mapping summary

Sample	Input reads	Uniquely mapped reads (%)	Multiple mapping reads (%)	Unmapped reads (%)	Chimeric reads (%)
S1	37251546	93.85	5.75	3.39	0
S2	37295556	94.01	5.57	0.41	0
S5	36097204	93.46	6.18	0.37	0
S6	37267910	93.82	5.72	0.46	0
S7	37295246	93.68	5.88	0.44	0
S8	37253286	93.77	5.84	0.39	0
S9	32242667	92.21	6.20	1.59	0
S10	37230618	92.96	5.97	1.06	0
S13	37233256	92.31	6.10	1.59	0
S14	37227233	92.72	6.06	1.22	0
S15	37226938	91.43	6.25	2.33	0
S16	37314408	92.04	6.04	1.93	0

c. Quality control of alignment

The reads mapping quality was examined with the package 'RSeQC' following the instructions.

- saturation of sequencing depth

```
junction_saturation.py -i sample.bam -r hg19.refseq.bed12 -o output
```

- reads coverage over gene body

```
geneBody_coverage.py -r hg19.housekeeping.bed -i /path/to/the directory containing BAM files -o output
```

- inner distance (or insert size) between two paired RNA reads

```
inner_distance.py -i sample.bam -o output -r hg19.refseq.b
```

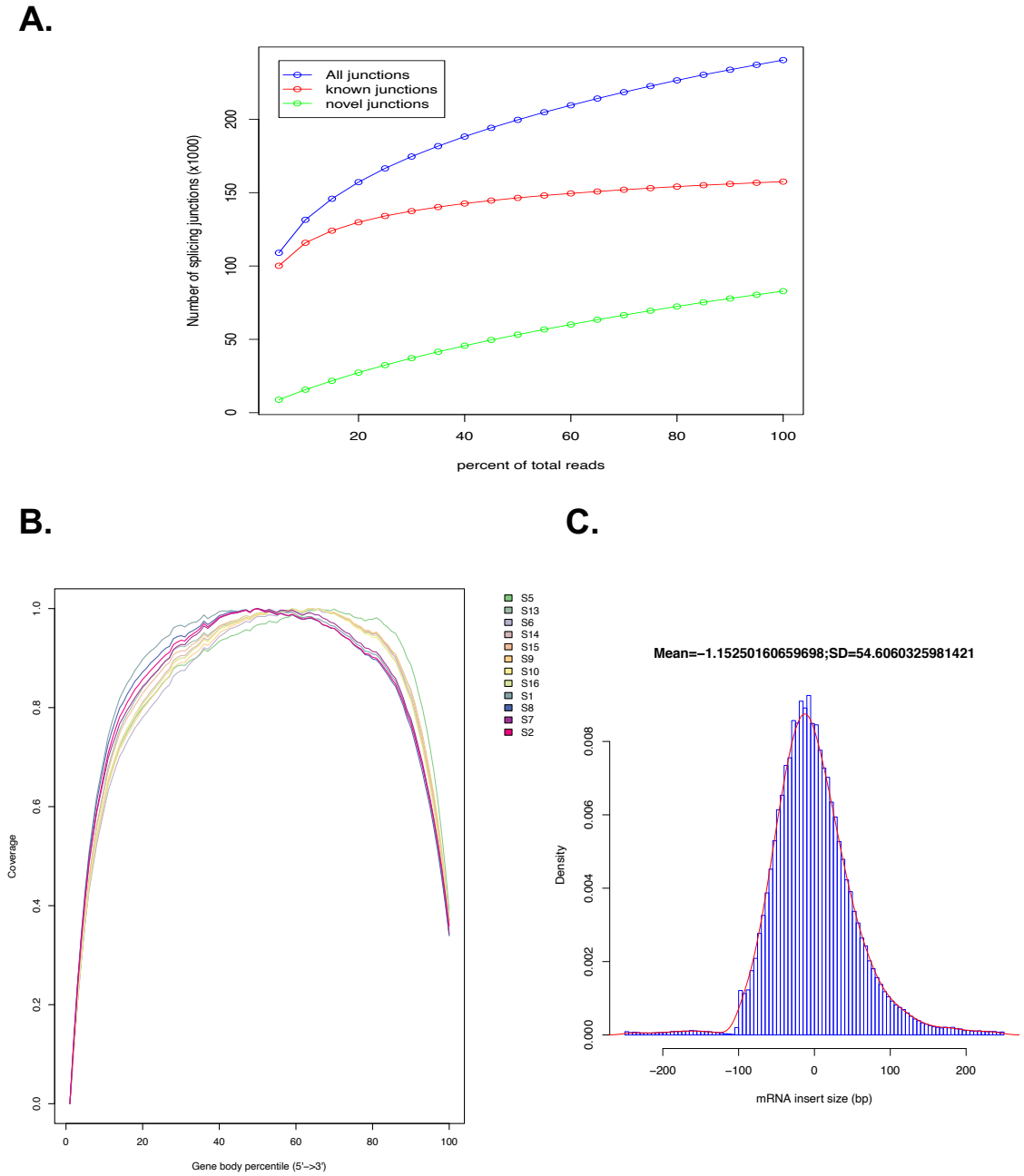

Fig. 11 Quality control of reads mapping. (**A**) Saturation analysis of sequencing depth for splice junctions. One representative output shows current sequencing depth is saturated for 'known junctions'

(red line) but not for 'novel junctions'. (**B**) The RNA-seq reads coverage over gene body. (**C**) One representative output showing the inner distance (or insert size) between two paired RNA reads.

- Reads distribution over genome feature

```
read_distribution.py -i sample.bam -r hg19.refseq.bed12
```

Table. 11 One representative output showing RNA-seq reads distribution over genome feature.

```
Total Reads                 74157576
Total Tags                  103633695
Total Assigned Tags         97730839
```

Group	Total_bases	Tag_count	Tags/Kb
CDS_Exons	35440972	73192762	2065.20
5'UTR_Exons	34573261	5607849	162.20
3'UTR_Exons	56295984	15264063	271.14
Introns	1444230997	3162355	2.19
TSS_up_1kb	30387574	59790	1.97
TSS_up_5kb	135838145	112870	0.83
TSS_up_10kb	243470833	149874	0.62
TES_down_1kb	32140566	146485	4.56
TES_down_5kb	139627575	287309	2.06
TES_down_10kb	246486380	353936	1.44

d. Counting reads

For reads quantitation, paired-end reads were managed with 'SAMtools' [227] and counted using 'featureCounts' in the 'Rsubread' package [228].

- Re-sorting the BAM files generated by the 'STAR' aligner

```
samtools sort sample.bam -o sample.sorted.bam
```

- Indexing sorted BAM files

```
samtools index sample.sorted.bam
```

- Running 'featureCounts'

```
SAMPLE_DIR="path/to/the file folder containing all sample.sorted.bam"
TOOL_DIR="path/to/folder containing featureCount package"
for file in $(ls $SAMPLE_DIR/*.sorted.bam)
do
    sample='basename $file'
    $TOOL_DIR/featureCounts -p -T 4 -s 2 -a /path/to/gencode.v19.annotation.gtf -t
exon -g gene_id -o /path/to/output/$sample"_featureCounts.txt" $file
done
```

e. Differential expression analysis

The expression quantification was done by the package 'DEseq2' [229] following the instructions. The p-values attained by the Wald test were corrected for multiple testing using the Benjamini and Hochberg method. The adjusted p-values (padj) were used to determine significant genes.

- Count matrix input

```
library("DESeq2")
dds <- DESeqDataSetFromMatrix(countData = cts,
                                    colData = coldata,
                                    design = ~ condition)
```

- Differential expression analysis

```
dds <- DESeq(dds)

res <- results(dds)
```

- Exporting results to CSV files

```
resOrdered <- res[order(res$pvalue),]

resSig <- subset(resOrdered, padj < 0.05)
write.csv(as.data.frame(resSig),
          file="resSig.csv")
```

- Results visualization

The differentially expressed genes were presented as volcano plots with the package 'EnhancedVolcano' following https://github.com/kevinblighe/EnhancedVolcano.

```
EnhancedVolcano(LTR8KO_resLFC,
                lab = rownames(LTR8KO_resLFC),
                x = 'log2FoldChange',
                y = 'padj',
                title = 'LTR8 KO versus Wild type',
                selectLab = c('CGA','CSH2',
'CSHL1','CYP11A1','CYP19A1','DEPDC1B','ERVW-
1','GDF15','GH2','HSD3B1','LGALS14','PSG9','PSG4','PSG3','PSG2','SDC1','ERVFRD-
1','CGB','LGALS16','LGALS3','NPNT','TEAD4','CNTNAP2','BIN1','CDX2','PODXL','HIF1A',
'PLAC8'),
                xlab = bquote(~Log[2]~ 'fold change'),
                pCutoff = 10e-6,
                FCcutoff = 2,
                pointSize = 4.0,
                labSize = 6.0,
                labCol = 'black',
                boxedLabels = TRUE,
                colAlpha = 1,
                shape = c(1, 4, 23, 25),
                cutoffLineType = 'twodash',
                cutoffLineWidth = 1.0,
                drawConnectors = TRUE,
                widthConnectors = 0.5,
                colConnectors = 'black',
```

```
border = 'full',
borderWidth = 1.0,
borderColour = 'black')
```

f. Gene ontology analysis

For LTR8 KO and control samples, the top 1000 differently expressed genes were selected for gene ontology analysis with the online package 'Gorilla' [230].

2.14 Transcriptome of human primary trophoblasts of EO-PE

This work was an extension of the previous study from our laboratory [219]. In this project, the sample size was increased to 8 controls and 10 PE samples. Human trophoblast sampling and the sequencing pipeline was performed following the protocol in [219]. The data analysis was done in collaboration with Dr. Amit Pande and Dr. Manvendra Singh.

Table. 12 Clinical characteristics of the trophoblast samples

Control_sample ID	Week of pregnancy	PE_sample ID	Week of pregnancy
Control 1	term (> 37)	PE 1	29
Control 2	term	PE 7	27
Control 3	term	PE 9	33
Control 4	term	PE 12	25
Control 5	term	PE 13	27
Control 6	term	PE 16	34
Control 10	term	PE 21	36
Control 15	term	PE 23	27
		PE 25	35
		PE 26	32

2.15 Trans-well invasion assay with SGHPL4 cells

The protocol for trans-well well invasion assay with SGHPL4 cells was modified from [231].

2.15.1 Matrigel preparation

a. Inserts, companion plates and pipet tips were prechilled to keep Matrigel in the liquid state.

b. Desired number of prechilled inserts were placed into a 24-well companion plate using a sterile tweezer.

c. 50 µl 0.1 mg/ml Matrigel (#356231 BD Biosciences, diluted with the prechilled serum free Ham's F10 medium) was added in each insert chamber (#3422 Corning, 8µm). Plates were incubated at 37 °C for ~ 4 hr.

2.15.2 Chemoattractant preparation in lower chamber

600 µl completed Ham's F10 medium was added to each lower chamber. The medium containing 50 ng/ml human Epidermal Growth Factor (hEGF, #AF-100-15 Peprotech) was used as the positive control and the FBS free medium was the negative control.

2.15.3 Cells preparation in upper chamber

a. SGHPL4 cells were treated with trypsin solution at 37 °C for ~ 3 min and harvested into the completed Ham's F-10media.

b. Cell suspension was centrifuged at 300 X *g* for 5 min and the supernatant was replaced with the FBS free Ham's F-10 media. Then wash was repeated once.

c. Cell concentration was counted and adjusted to ~ 1 X 10^6 cells/ml with the FBS free Ham's F-10 media.

d. 100 µl cell suspension (~ 1 x 10^5 cells) was added to each trans-well insert and plates were incubated at 37 °C for 16 hr.

2.15.4 Fixation and staining

a. Cell medium was removed and trans-well inserts were washed twice in PBS.

b. 3.7% PFA in PBS was added and incubated for 5 min at RT for cell fixation. Then trans-well inserted were washed twice in PBS.

c. 100% methanol was added and incubated for 20 min at RT for permeabilization. Then trans-well inserted were washed twice in PBS.

d. 0.2% crystal violet were added and incubated for 30 min at RT for staining in dark. Then trans-well chambers were washed twice in PBS.

e. Non-invaded cells were gently scraped off with cotton swabs.

g. The invasive cells at four corners of each insert chamber were captured under a light microscope (10X).

2.15.5 Quantification

Invaded cells were counted by the '`Fiji`' package (https://imagej.net/Fiji) with the following scripts.

```
run ("Invert LUT");
setAutoThreshold("Default dark");
//run("Threshold...");
setThreshold(0, 45);
//setThreshold(0, 45);
setOption("BlackBackground", true);
run("Convert to Mask");
run("Analyze Particles...", "size=100-Infinity show=Outlines clear summarize");
```

2.16 Statistical analysis

The statistics in the RNA-seq analysis was done following the default methods in the packages. Other statistical analyses were performed using GraphPad Prism 9 (GraphPad Software, San Diego, California USA, www.graphpad.com). The data are representative of mean ± SD and *P* values less than 0.05 were considered statistically significant.

2.17 Primers or oligos used in this study

Table. 13 Primers or oligos used in this project

Usage	Target	Sequence (5‘-> 3‘)
RT-qPCR	*β-ACTIN*	Forward: CACCAACTGGGACGACAT
		Reverse: ACAGCCTGGATAGCAACG
	18rRNA	Forward: ACATCCAAGGAAGGCAGCAG
		Reverse: TTTTCGTCACTACCTCCCCG
	GATA2	Forward: AGTCTGTCTATTGCCTGCCG
		Reverse: CGCGGGTTCCATGTAGTTGT
	GATA3	Forward: GAAGGCAGGGAGTGTGTGAA
		Reverse: GTCTGACAGTTCGCACAGGA
	TFAP2A	Forward: GAGAACAGAACAGGCCGTGA
		Reverse: TGACTCAGTCCCATGAAGCG
	TFAP2C	Forward: AGAGCCTTCACTGGTTCTGC
		Reverse: GCGGACACAAAAACCAACCA
	CGB	Forward: TACTGCCCCACCATGACC
		Reverse: CACGGCGTAGGAGACCAC
	HERVW-1	Forward: CCCCATCGTATAGGAGTCTT
		Reverse: CCCCATCAGACATACCAGTT
	HERVFRD-1	Forward: CCACAACCAATTCCGCCATC
		Reverse: ATAATCAGCAGGCCGGAACC
	PSG9_all transcripts	Forward: TCTGGATCCCAGGCTCATCT
		Reverse: GGGCGGGTTCCAGAAGTTTA
	PSG9_subgroup B transcripts	Forward: TCTCAGTCATGACTGCAACAAC
		Reverse: GGGCATTCAGATAGACAGCAA
	PSG9_subgroup C transcripts	Forward: AAGCTCACACTTTTCCCCCA
		Reverse: GCCCTTTCTACACACACGCT
	PSG1	Forward: TGAGAATTGCTCCTGCCCTG
		Reverse: CGGAAACTTTGGTTGGCTCG
	PSG5	Forward: AAGCATAGAGGGCTCTATAC
		Reverse: AGGAGAGGAAGACGTCCTAT
	GATA2 (CCDS 3049.1)	Forward: EcoR V - acgtcccagactacgctgatatcATGGAGGTGGCGCCCGAG
		Reverse: aataaacaagttaacaacgcggccgcCTAGCCCATGGCGGTCACC - Not I

Overexpression of gene of interest (GOI)	*GATA3* (CCDS 31143.1)	Forward: EcoR V - acgtcccagactacgctgatatcATGGAGGTGACGGCGGAC
		Reverse: aataaacaagttaacaacgcggccgcCTAACCCATGGCGGTGAC - Not I
	TFAP2A (CCDS 4510.1)	Forward: EcoR V - acgtcccagactacgctgatatcATGCTTTGGAAATTGACG
		Reverse: aataaacaagttaacaacgcggccgcTCACTTTCTGTGCTTCTC - Not I
	TFAP2C (CCDS 13454.1)	Forward: EcoR V - acgtcccagactacgctgatatcATGTTGTGGAAAATAACCG
		Reverse: aataaacaagttaacaacgcggccgcTTATTTCCTGTGTTTCTCC - Not I
	PSG9 isoform -1 (UniProt Q00887-1)	Forward: EcoR V - acgtcccagactacgctgatatcATGGGGCCCCTCCCAGCC
		Reverse: aataaacaagttaacaacgcggccgcTCATGACTGAGACTCTGTCAGGTCTCC - Not I
	PSG9 isoform -2 (UniProt Q00887-2)	Forward: EcoR V - acgtcccagactacgctgatatcATGGGGCCCCTCCCAGCC
		Reverse: aataaacaagttaacaacgcggccgcTCAGGGCTGATAAAGCCCCCTC - Not I
Plasmids for ployadenylation reporter assay	full-length (3'UTR + MER65-int) (NM_002784.5)	Forward: Xho I - cagtaattctaggcgatcgctcgagCTGCAACAACTGAGACACTG
		Reverse: ttagggggggggagggagaggggcgaattcCTGTTAACTTACTCTCAGAGGC - EcoR I
	3' UTR	Forward: Xho I - cagtaattctaggcgatcgctcgagCTGCAACAACTGAGACACTG
		Reverse: ttagggggggggagggagaggggcgaattcTTACGGAACATGTGCAAATAAC - EcoR I
	MER65-int	Forward: Xho I - cagtaattctaggcgatcgctcgagACAAAGAAGAAAAAAACTCAATG
		Reverse: ttagggggggggagggagaggggcgaattcCTGTTAACTTACTCTCAGAGGC - EcoR I
sgRNA for LTR8 KO	the LTR8 element at PSG9 locus	sgRNA1_forward: CACCGCTCTAACTAGGCTGGCTGTT
		sgRNA1_reverse: AAACAACAGCCAGCCTAGTTAGAGC
		sgRNA2_forward: CACCGTTGAAATGGTGGTCTGTAGT
		sgRNA2_reverse: AAACACTACAGACCACCATTTCAAC
		sgRNA3_forward: CACCGCTCTGACACCCTGGTTAGTC
		sgRNA3_reverse: AAACGACTAACCAGGGTGTCAGAGC
Genotyping for LTR8 KO	the LTR8 element at PSG9 locus	Forward: GGCATCCAAACCTCAGAGGAG
		Reverse: TTAGGAAAAATGGGGAGGACTGCAA

2.18 Antibodies used in this study

Table. 14 Antibodies used in this project

Usage	1st antibody	2nd antibody
Western blot	anti-HA, 50 ng/mL, #11867423001 Roche	Goat anti-Rat IgG (H+L) HRP, 1:5,000 #31470, Thermo Fisher Scientific
	anti-FLAG, 2 µg/mL, #F3165 Sigma-Aldrich	Goat anti-Mouse IgG (H+L) HRP, 1:5,000 #31432, Thermo Fisher Scientific
Membrane staining	anti-HA, 5 µg/mL, #11867423001 Roche	Donkey anti-Rabbit IgG (H+L) 488, 2 µg/mL # A-21206 Thermo Fisher Scientific
	anti-E-Cadherin, 10 µg/mL, #AF748 R&D System	Donkey anti-Goat IgG (H+L) 647, 2 µg/mL # A-21206 Thermo Fisher Scientific
Intracellular staining	anti-HA, 5 µg/mL, #11867423001 Roche	Goat anti-Rat IgG (H+L) 488, 2 µg/mL # A-11006 Thermo Fisher Scientific
	anti-Calnexin, 1:100, #2679 Cell Signaling Technology	Goat anti-Rabbit IgG (H+L) 647, 2 µg/mL # A-21244 Thermo Fisher Scientific

3. Results

3.1 The LTR8 promotes PSG9 transcription as a placenta-specific enhancer

3.1.1 The LTR8 at *PSG* gene cluster exhibits the chromatin features associated with enhancer activity

Given the emerging roles of HERVs in gene regulation, we asked whether HERV-derived elements contribute to PSG expression. To this end, we looked into the annotated HERVs at *PSG* gene cluster and screened the candidates with potentially modulatory bioactivity based on the surrounding chromatin characteristic in human trophoblasts. Two LTR-derived elements, the LTR8 and MER65-int, were identified to be present across the entire *PSG* gene cluster (Fig. 12A). Dramatically, the LTR8, a solitary element from HERV-K family, is located at the second intron of each *PSG* gene and enriched with the enhancer-associated histone modifications in the human pluripotent stem cells (hPSCs)-derived trophoblast progenitors (TBs)[24], including the H3K4me1 and H3K27ac (Fig. 12). In addition, during the SCT differentiation from the primary VCTs in *vitro* [232], the LTR8 seems to exhibit the enrichment of H3K4me3 in the SCT-like cells (Fig. 12). Furthermore, the chromatin accessibility analysis with the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) [72] showed the local chromatin at LTR8 locus exhibits relatively open state in the TBs (Fig. 12). Importantly, mining the online Chromatin Immunoprecipitation Sequencing data (CHIP-seq) from the hPSCs-derived TBs [24], we found LTR8 contains the predictive binding sites of trophoblast specification factors (TSFs), such as GATA3 and TFAP2A (Fig. 12). Altogether, these chromatin features suggested the LTR8 could be a potential enhancer for PSG transcription. (The data mining was done in collaboration with Dr. Manvendra Singh)

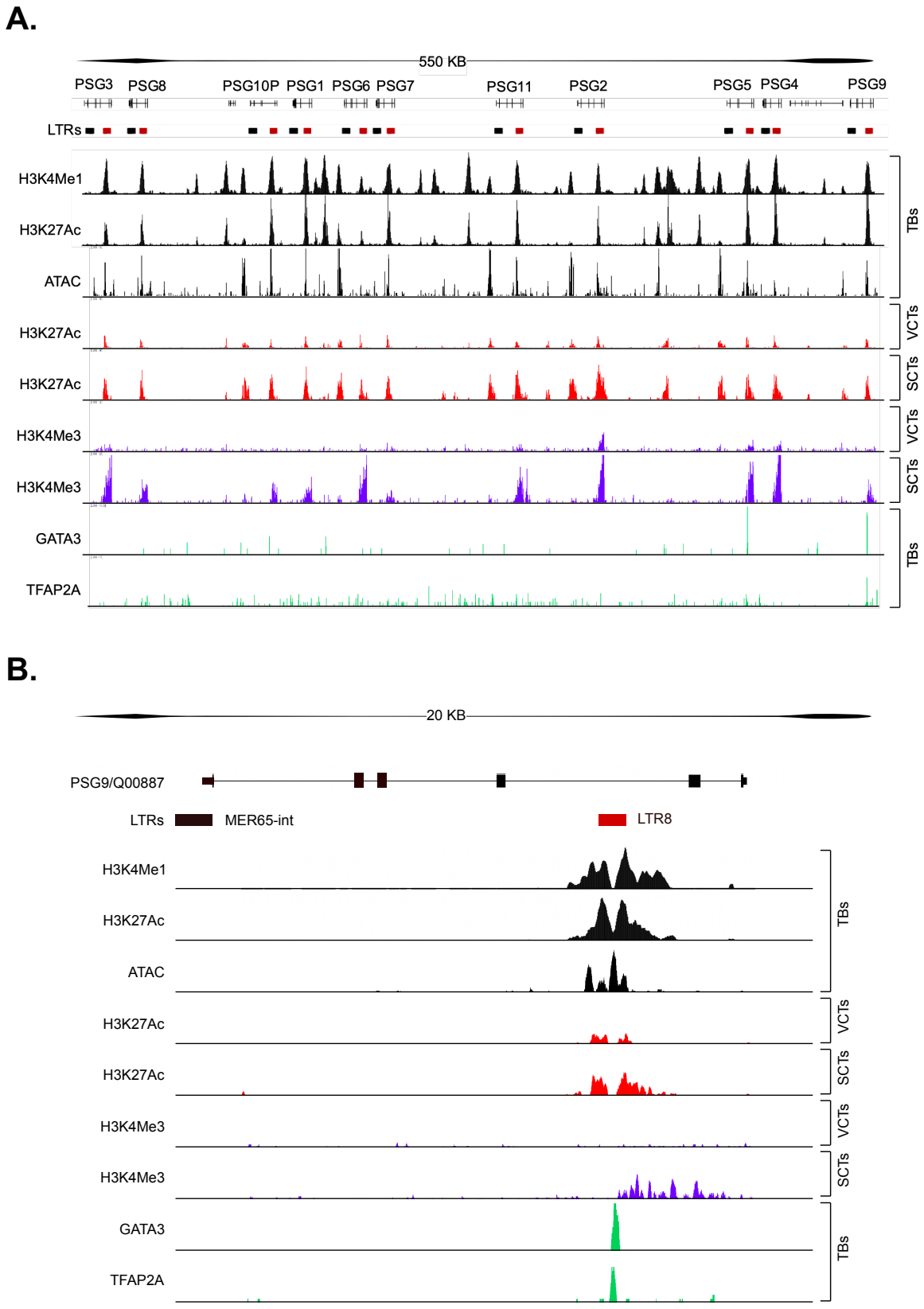

Fig. 12 Genome browser snapshots showing the chromatin features at human *PSG* gene cluster. (**A**) Two LTR elements, LTR8 and MER65-int, are ubiquitous across the entire *PSG* gene cluster. The

chromatin features, including histone modifications, chromatin accessibility and the transcription factor binding signals, were mined from the online available resource. (**B**) An overview of the chromatin landscape at *PSG9* gene locus. TBs, the trophoblast progenitors derived from hPSCs; VCTs, villus cytotrophoblasts; SCTs, syncytiatrophoblasts. Snapshots were modified from that made by Dr. Manvendra Singh.

3.1.2 The LTR8 is required for PSG9 transcription as an enhancer

To validate the enhancer activity of LTR8 experimentally, we first performed the luciferase reporter assay with other candidate elements among 8kb upstream of *PSG9* transcription starting site (TTS) in trophoblast cells (Fig. 13A and B). Significantly, the LTR8 displayed the enhancer activity by exerting robust luminescence signals compared with that of negative control and other surrounding elements (Fig. 13C). To further examine the regulatory role of LTR8 in endogenous PSG9 expression, we carried out the CRISPR-directed LTR8 deletion with two kinds of guide RNAs pairs flanking the LTR8 copy at *PSG9* locus (Fig. 14A, B and C). As expected, the mRNA expression of PSG9, but not that of other PSGs, was dramatically decreased upon the loss of LTR8 in all homozygous colonies (Fig. 14D and E), indicating an essential role of the LTR8 copy in PSG9 transcription. It is worth noting that the downregulation effect of LTR8 deletion on *PSG9* mRNA expression seemed to become weak after six passages cell culture (data not shown), meaning the LTR8 enhancer activity for PSG9 expression might be gradually compensated by unappreciated mechanisms.

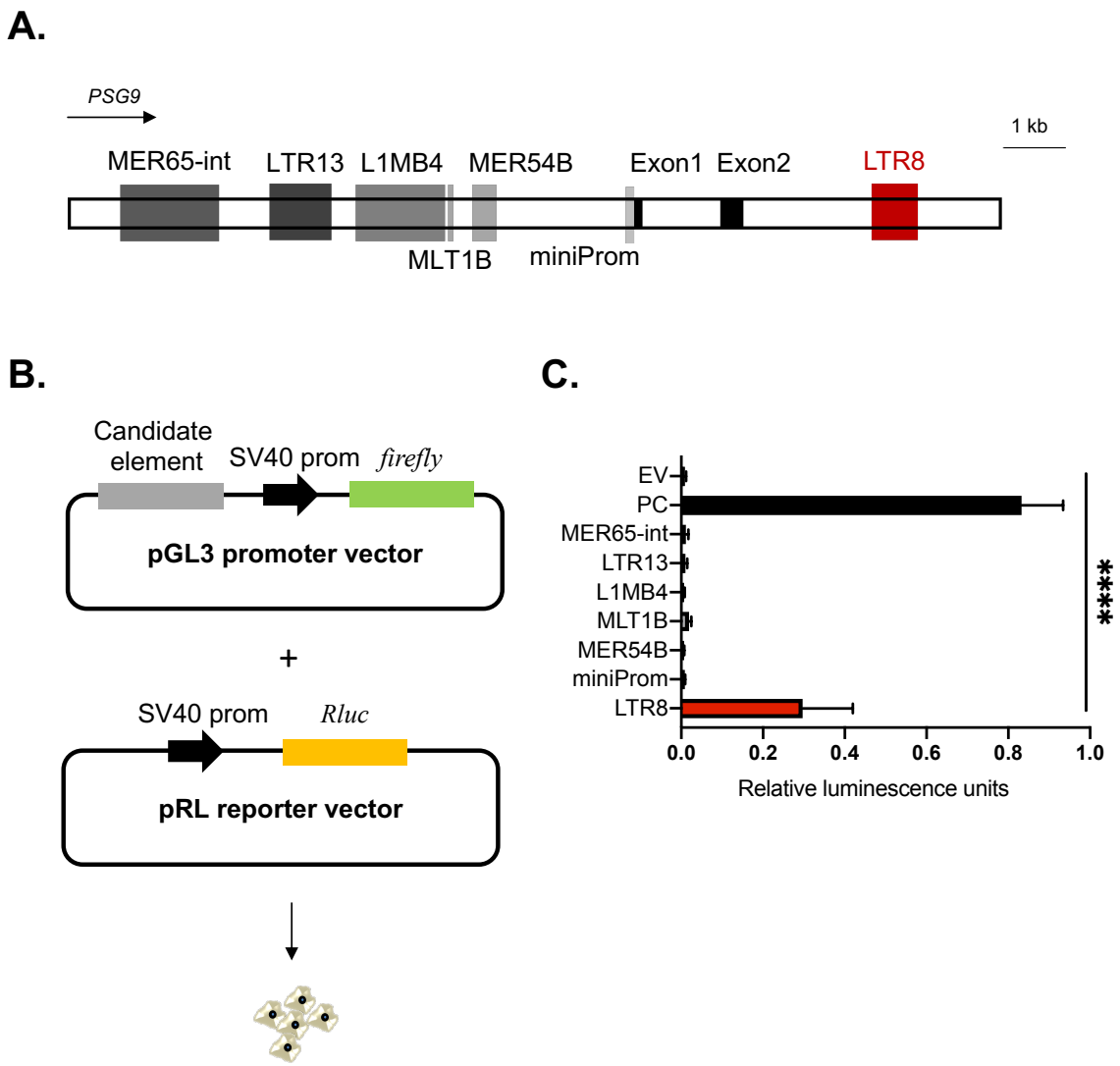

Fig. 13 Enhancer luciferase assay of candidate elements at *PSG9* gene locus. (**A**) Schematic diagram showing the candidate elements at *PSG9* locus for the enhancer luciferase assay. (**B**) Diagram depicting the strategy for the enhancer luciferase assay (see details in the methods). Basically, the pGL3 promoter vector containing the candidate element and the internal control vector were co-transfected in trophoblasts. The enhancer activity was evaluated based on the luminescence output. Plasmids were constructed by Dr. Julianna Rugor. (**C**) One representative result of three independent replicates. N = 6 technical replicates at least, mean ± SD. Ordinary one-way ANOVA followed by multiple comparation. EV vs. PC or LTR8, **** $P < 0.0001$; EV vs. any other element from the rest, not significant. EV, the empty pGL3 promoter vector as a negative control; PC, the pGL3 promoter vector containing a SV40 enhancer as a positive control.

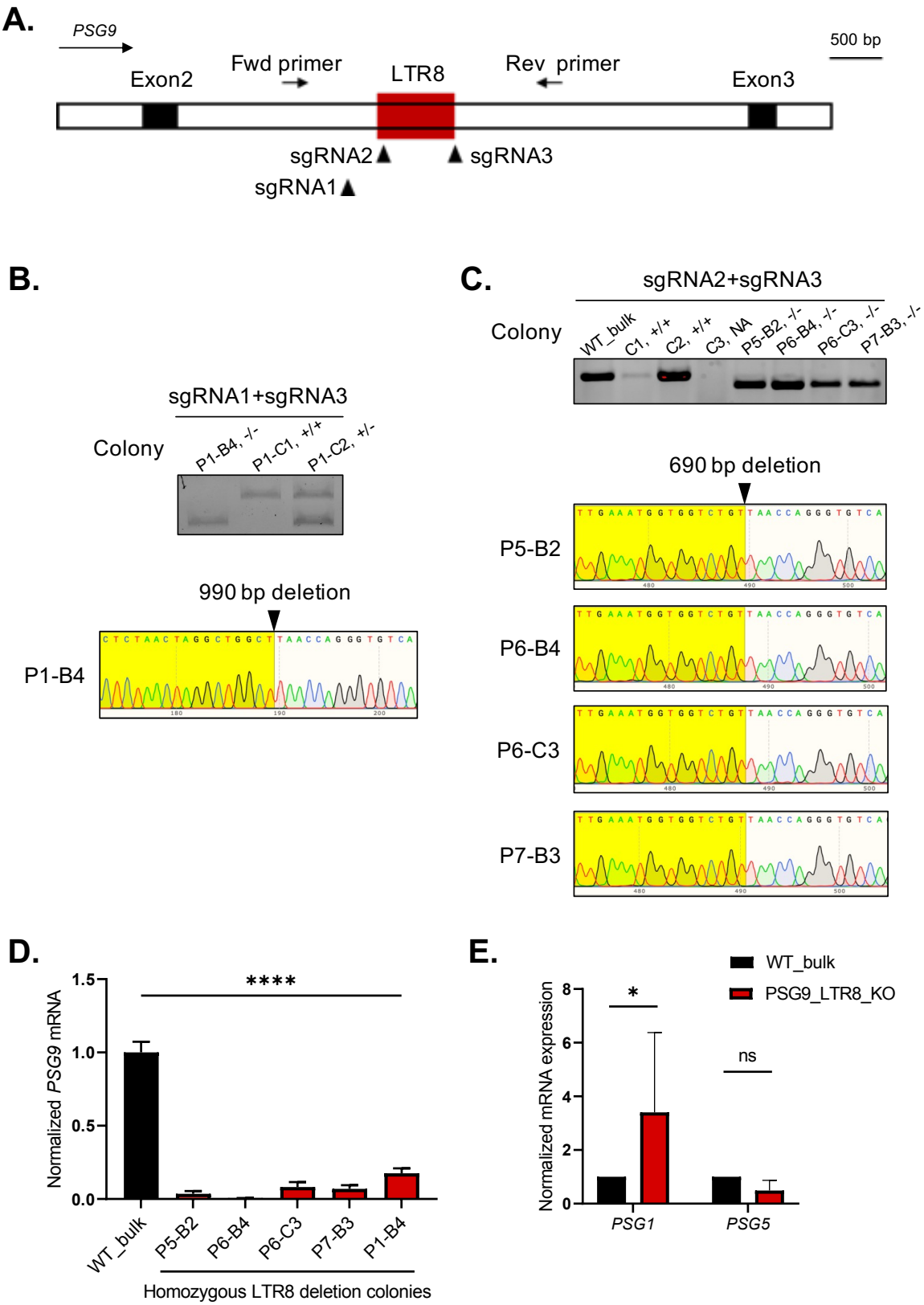

Fig. 14 CRISPR-directed LTR8 deletion at *PSG9* locus. (**A**) Schematic diagram showing the sgRNAs targeting the LTR8 at the sencond intron of *PSG9* gene. (**B**) The CRISPR-directed LTR8 deletion (990bp) with the guide RNA pair, shRNA1 and sgRNA3, was confirmed by genotyping and Sanger sequencing. (**C**) The CRISPR-directed LTR8 deletion (690bp) with the guide RNA pair,

shRNA2 and sgRNA3, was confirmed by genotyping and Sanger sequencing. (**D**) The normalized mRNA expression of PSG9 in the homozygous colonies of LTR8 deletion. N = 3 technical replicates, mean ± SD. Ordinary one-way ANOVA followed by multiple comparation. Wide type bulk cell population vs. each homozygous LTR8 deletion colony: P5-B2, P6-B4, P6-C3, P7-B3 or P1-B4, **** $P < 0.0001$. (**E**) The normalized mRNA expression of PSG1 and PSG5 in the LTR8 deletion background. The average value of mRNA expression in five homozygous colonies of LTR8 deletion was displayed as mean ± SD. Unpaired t test. * $P = 0.044$; ns, not significant.

3.1.3 The LTR8 element promotes PSG9 transcription as an enhancer bound by GATA2 and GATA3

Given that the LTR8 enhancer contains the potential binding sites of TSFs (Fig. 12B), we interrogated the regulatory role of the TSFs in LTR8-directed PSG9 transcription. Indeed, ectopic expression of TSFs, such as GATA3 and TFAP2A, can improve the enhancer performance of LTR8 in the luciferase reporter assay (Fig. 15A). However, PSG9 mRNA expression was only upregulated upon overexpression of either GATA2 or GATA3, indicating GATA2 and GATA3 could promote the PSG9 transcription in the endogenous chromatin context (Fig. 15B). Next, to investigate whether the LTR8 is required for this process, we overexpressed GATA3 in the LTR8 knockout cells (LTR8 KO) and examined PSG9 expression. As shown in the Fig. 15C, the upregulation of PSG9 mRNA was not compromised upon GATA3 overexpression in LTR8 deletion background compared with that in wide type control, suggesting the LTR8 might be sufficient but not essential for the ectopic GATA3-induced PSG9 transcription. Interestingly, despite the potential interaction between TSFs and the LTR8 copy at *PSG1* gene locus, no significant change in *PSG1* mRNA expression was detected upon TSFs overexpression, which means different PSG members might be divergently regulated (Fig. 15D). Altogether, these data indicate that LTR8 could control PSG9 transcription as an enhancer with binding to GATA2 and GATA3.

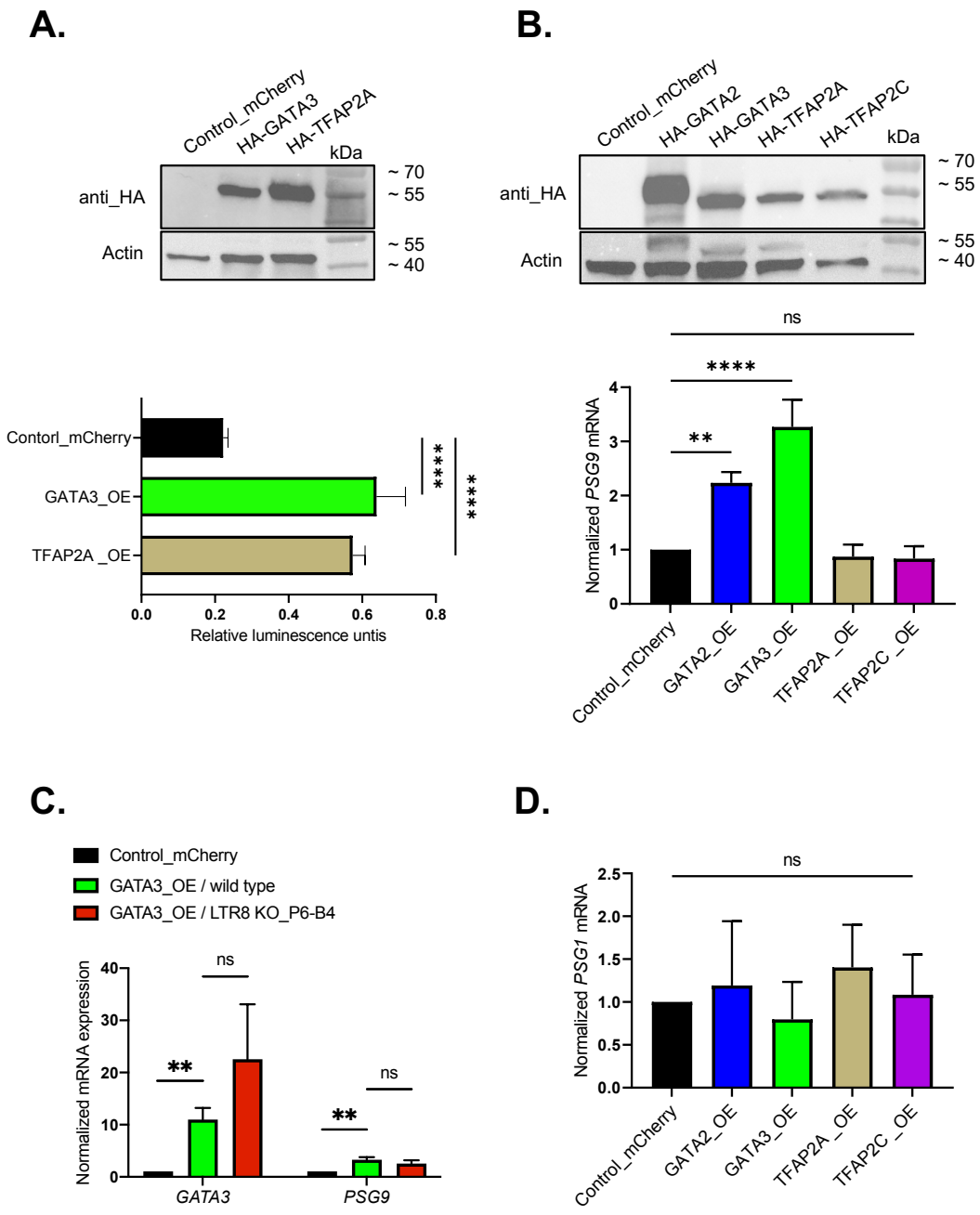

Fig. 15 GATA2 and GATA3 could promote LTR8-directed *PSG9* mRNA expression. (**A**) The luciferase assay to measure the enhancer activity of the LTR8 in TSF overexpression background. The expression of ectopic TSFs was validated by weastern blot. One representative result of three independent replicates. N = 8 technical replicates, mean ± SD. Ordinary one-way ANOVA followed by multiple comparation. Control vs. GATA3 or TFAP2A, **** $P < 0.0001$. (**B**) The mRNA expression of PSG9 in TSF overexpression background. The expression of ectopic TSFs was validated by weastern blot. N = 3 independent replicates, mean ± SD. Ordinary one-way ANOVA followed by multiple comparation. Control vs. GATA2 OE, ** $P = 0.0011$; control vs. GATA3 OE, **** $P < 0.0001$; control

vs. TFAP2A or TFAP2C, not significant (ns). (**C**) The mRNA expression of PSG9 in LTR8 deletion background with GATA3 overexpression. N = 2 independent replicates at least, mean ± SD. Unpaired t test. ** $P = 0.004$ for GATA3 mRNA expression and ** $P = 0.003$ for PSG9 mRNA expression; ns, not significant. (**D**) The mRNA expression of PSG1 in TSF overexpression background. N = 3 independent replicates, mean ± SD. Ordinary one-way ANOVA followed by multiple comparation. Control vs. each TSF overexpression, not significant (ns). OE, overexpression.

3.1.4 The LTR8 could function as a placenta-specific enhancer in anthropoid primates

The comparative genomic alignment revealed that the two LTR elements at the *PSG9* locus, LTR8 and MER65-int, are present in anthropoid primates but neither in prosimians nor non-primate mammals (Fig. 16A). Importantly, mining the recently published ATAC-seq dataset showing the chromatin accessibility profiles from 59 human fetal samples at single-cell levels [233], we found the LTR8 chromatin at *PSG9* locus is particularly accessible in placental cell types, including trophoblasts and trophoblast giant cells (Fig. 16B), suggesting the LTR8 might function as a placenta-specific enhancer in anthropoid primates. Furthermore, the High throughput chromosome conformation Capture (Hi-C) profiling of the H1 hESC-derived trophectoderm cells [234,235] showed human *PSGs* and the closely related *CEACAM* gene family share the same topologically associating domain (TAD, the basic unit of 3D genome structure) (Fig. 17A), where the LTR8 at *PSG9* locus might potentially interact with the two genomic regions at chromosome 19 (GRCh38/hg38, distance normalized interaction frequency >2, interaction range = 2Mb) (Fig. 17A). In addition, the Promoter Capture Hi-C (PC Hi-C) in human trophoblasts [234,236] showed that the *PSG9* promoter could interact with that of *PSG4* and other genomic regions (GRCh37/hg19, *p*-value < -log10(2), interaction range = 1Mb) (Fig. 17B). Taken together, those data suggested the LTR8 could control PSG9 transcription as a placenta-specific enhancer in anthropoid primates. However, given the potential interaction between the LTR8 and other genomic regions, whether the LTR8 is a specific enhancer for PSG9 expression needs to be further deciphered.

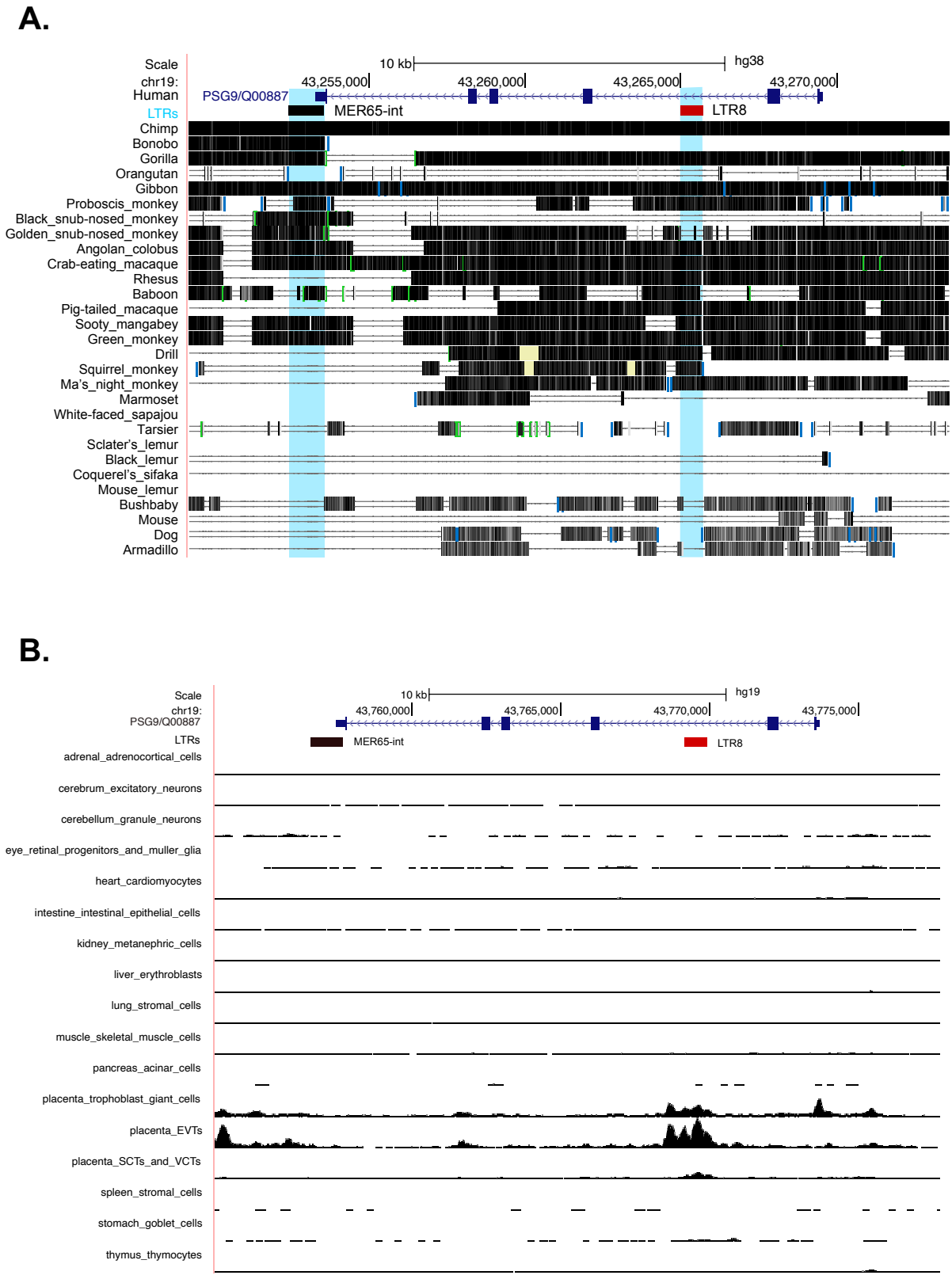

Fig. 16 Genome browser snapshot showing the species- and tissue-specifity of the LTR elements. (**A**) Comparative genomics (GRCh38/hg38) showing that the LTR8 and MER65-int at *PSG9* locus are especially present in anthropoid primates. (**B**) Genome browser snapshot (GRCh37/hg19) showing the chromatin accessibility of LTR8 in multiple human fetal cell types.

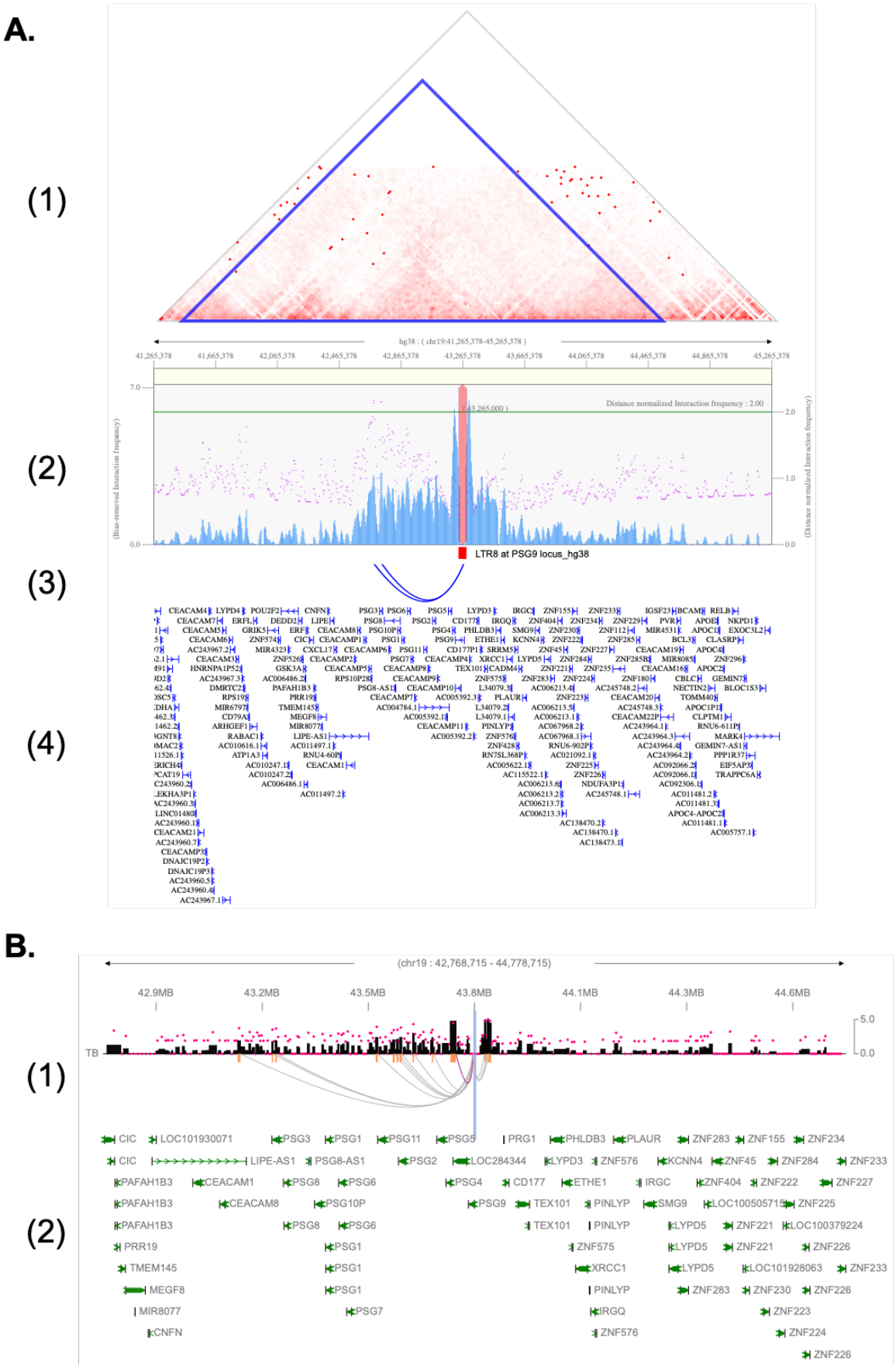

Fig. 17 Chromatin interaction profiling of the *PSG9* gene region. (**A**) Hi-C data visualization showing the chromatin interaction of the LTR8 element in H1 ESCs-derived

trophectoderm cells (GRCh38/hg38). （**1**） shows human *PSGs* and the closely related *CEACAM* genes belong to the same TAD indicated in blue. (**2**) depicts the one-to-all interaction frequency. Blue bar graph represents bias-removed chromatin interaction frequency, and magenta dots represent distance-normalized interaction frequency. (**3**) shows the arc-diagram of the identified interactions with the distance normalized interaction frequency >2, interaction range = 2Mb. (**4**) shows the genes among the interaction range. (**B**) The PC Hi-C shows the interaction of *PSG9* promoter with *PSG4* promoter and other genomic regions (GRCh37/hg19). (**1**) shows the arc-diagram of the identified interactions with the *P*-value < -log10 (2), interaction range = 1Mb. (**2**) shows the genes among the interaction range. Data mining using the online resource from [235,236] and visualization by the online web tool '3D-Genome Interaction Viewer & database (3DIV)' (http://www.3div.kr/hic and http://www.3div.kr/capture_hic) [234].

3.2 The MER65-int is involved in alternative polyadenylation and isoform formation of PSG9

3.2.1 The MER65-int contributes to mRNA polyadenylation of PSG9 through providing *cis*-elements and polyadenylation signals

The MER65-int element is located at the 3' end of each *PSG* gene member. It has been notorious that PSGs and the closely related CEACAMs are highly divergent at least partially because of the extensive alternative splicing and polyadenylation variants [1,108,118]. Despite the complexity, we found *PSG9* transcripts could be divided into three subgroups, named A, B, C for simplicity, based on the 3' terminus merges with the MER65-int element or not (Fig. 18A). For instance, in the subgroup A and B, the MER65-int merges with the 3' untranslated region (3'UTR), while it is located ~ 3.5kb downstream of the 3'UTR in the transcripts of subgroup C (Fig. 18A). Given the proposed role of ERVs in mRNA polyadenylation [214,215], we hypothesized that the MER65-int could contribute to the alternative polyadenylation (APA) of *PSG9* mRNA, in which multiple polyadenylation sites (PASs) at pre-mRNA are recognized by cleavage and polyadenylation machinery, leading to transcript isoforms with various 3′ UTRs [237,238]. Firstly, to examine the polyadenylation efficiency of MER65-int, we used a recently developed bicistronic reporter to measure the strength of candidate elements in signaling transcription pause [220] (Fig. 18B). As shown in Figure 18C, the strength of solitary 3'UTR in transcription termination was significantly compromised

compared to that of the full-length fragment, whereas the MER65-int alone outperformed the solitary 3'UTR and achieved nearly full pause effect, suggesting the downstream part of the MER65-int was required for an efficient mRNA polyadenylation.

Besides the copy at *PSG9* locus, we found MER65-int merges the 3'UTR of other *PSG* transcripts, suggesting the MER65-int-directed mRNA polyadenylation could be applicable to the entire human *PSG* family (*PSG1-PSG9*, *PSG11*). Consistently, compared to their counterpart alignments in the consensus sequence, the MER65-int copies fused with the 3'UTR exhibit higher possibility to harbor the core upstream *cis*-element (AAUAAA) for mRNA polyadenylation [239] (Fig. 18D), suggesting the MER65-int-directed mRNA polyadenylation might be specific for *PSG* genes in human genome. Taken together, these data indicated that MER65-int could contribute to the mRNA polyadenylation of *PSG* genes by providing *cis*-elements and/or PASs.

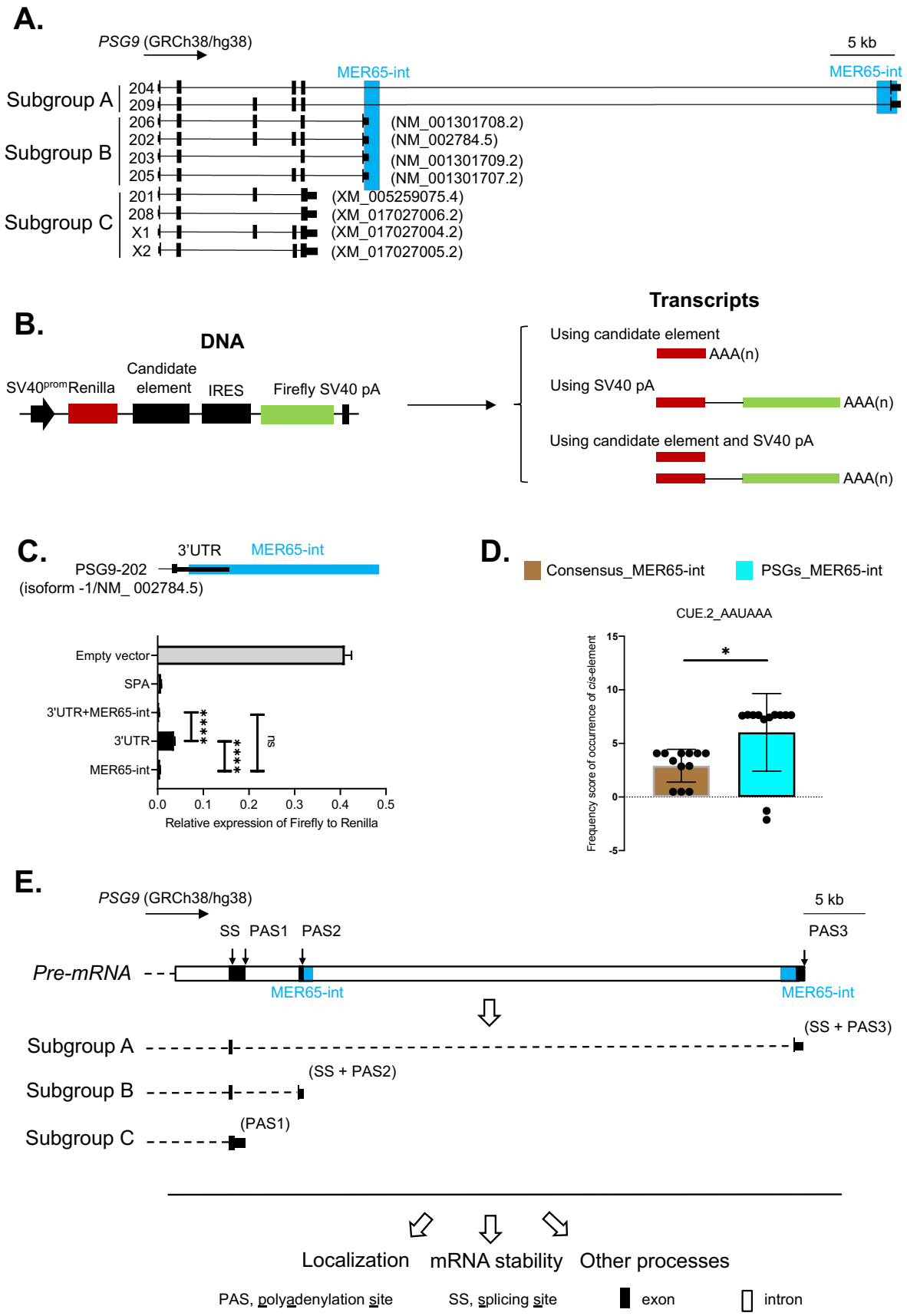

Fig. 18 The MER65-int is involved in mRNA polyadenylation of *PSG* genes. (**A**) Diagram depicting that *PSG9* transcripts can be classified into three subgroups based on whether the 3'terminus merges

with the MER65-int or not (GRCh38/hg38). The Ensembl transcript names were shown on the left terminus and NCBI reference numbers were shown on the right terminus. The last two transcripts were named as X1 and X2, respectively, in this study. (**B**) Diagram depicting the working principle of the luciferase reporter assay to measure the polyadenylation efficiency of candidate elements [220]. The candidate element was introduced into a bicistronic reporter which can be transcribed into the bicistronic mRNA with two open reading frames (ORFs), the *Renilla* and *Firefly*. Polyadenylation strength of candidate element is expected to be evaluated based on the relative expression ratio of Firefly to Renilla. See details in methods. (**C**) One representative result of three independent replicates. N = 8 technical replicates, mean ± SD. Ordinary one-way ANOVA followed by multiple comparation. The full-length (3'UTR + MER65-int) vs. solitary 3'UTR, **** $P < 0.0001$; 3'UTR vs. MER65-int, **** $P < 0.0001$; the full-length vs. MER65-int, not significant (ns). EV, the empty vector without the candidate element as a negative control; SPA, a known synthetic polyadenyaltion signal as a positive control. (**D**) Evaluation of occurrence of the core upstream *cis*-element (CUE.2_AAUAAA) for polyadenylation in MER65-int sequences, see details in the methods. Paired t test, * $P = 0.022$. Each dot represents one MER65-int sequence in *PSG* locus or the corresponding alignment in the consensus sequence. (**E**) A proposed model showing the MER65-int-directed alternative polyadenylation acts in concert with the alternative splicing procedure to contribute to isoform formation of PSG9.

3.2.2 MER65-int - directed alternative polyadenylation leads to the PSG9 isoforms with different locations

As mentioned above, the subgroup C transcripts of *PSG9* specially contains the 3' terminus isolated from the MER65-int. Next, we asked whether the PSG9 isoforms can be generated from MER65-int-directed APA (Fig. 18E). Indeed, compared to those from subgroup A and B, the transcripts from subgroup C display a dominant variance in the last exons probably resulting from the interaction between the alternative splicing and the APA procedure. In addition, all 4 annotated transcripts of subgroup C can be confirmed by RT-PCR and sequencing (Fig. 19A). For further characterization in detail, we selected the isoform -1 (UniProt, Q00887-1) and isoform -2 (UniProt, Q00887-2) as the representative members from the subgroup B and C, respectively (Fig. 19B). PSG9 isoform - 1 and isoform - 2 share 64.6% amino acid identity. Dramatically, compared with the canonical structure of isoform -1 (L/N-A1-A2-B2) with a short hydrophilic C-terminus, the isoform -2 is lake of the A2 domain but contains a long hydrophobic C-terminus with a predicted transmembrane region (Fig. 19C and D). All human PSGs have been recognized to be secreted except the

member, termed PSG11w, which was reported to be an intracellular protein and retained at the endoplasmic reticulum (ER) [123]. Interestingly, the PSG9 isoform -2 and the PSG11w share the identical C-terminus and may be the same protein from the subgroup C. We speculated the confusion could be due to the inconsistent PSG nomenclature over time [2]. Whatever the reason, to confirm the localization of PSG9 isoforms, we engineered the isoform -1 and isoform -2 fused with different tag proteins and co-expressed them into trophoblast cells (Fig. 19E). Of note, the cell fractionation analysis showed that the FLAG-tagged PSG9 isoform -1 was secretory and detectable in culture media, whereas the HA-fused isoform -2 was enriched in the membrane compartment (Fig. 19F). Consistent with the published data [123], our immunostaining results further validated the membrane bound isoform -2 was intracellular and co-localized with an ER maker protein (Fig. 19G and H). Altogether, our experimental data demonstrated that the MER65-int-directed mRNA polyadenylation contributes to PSG9 isoform formation with different localizations.

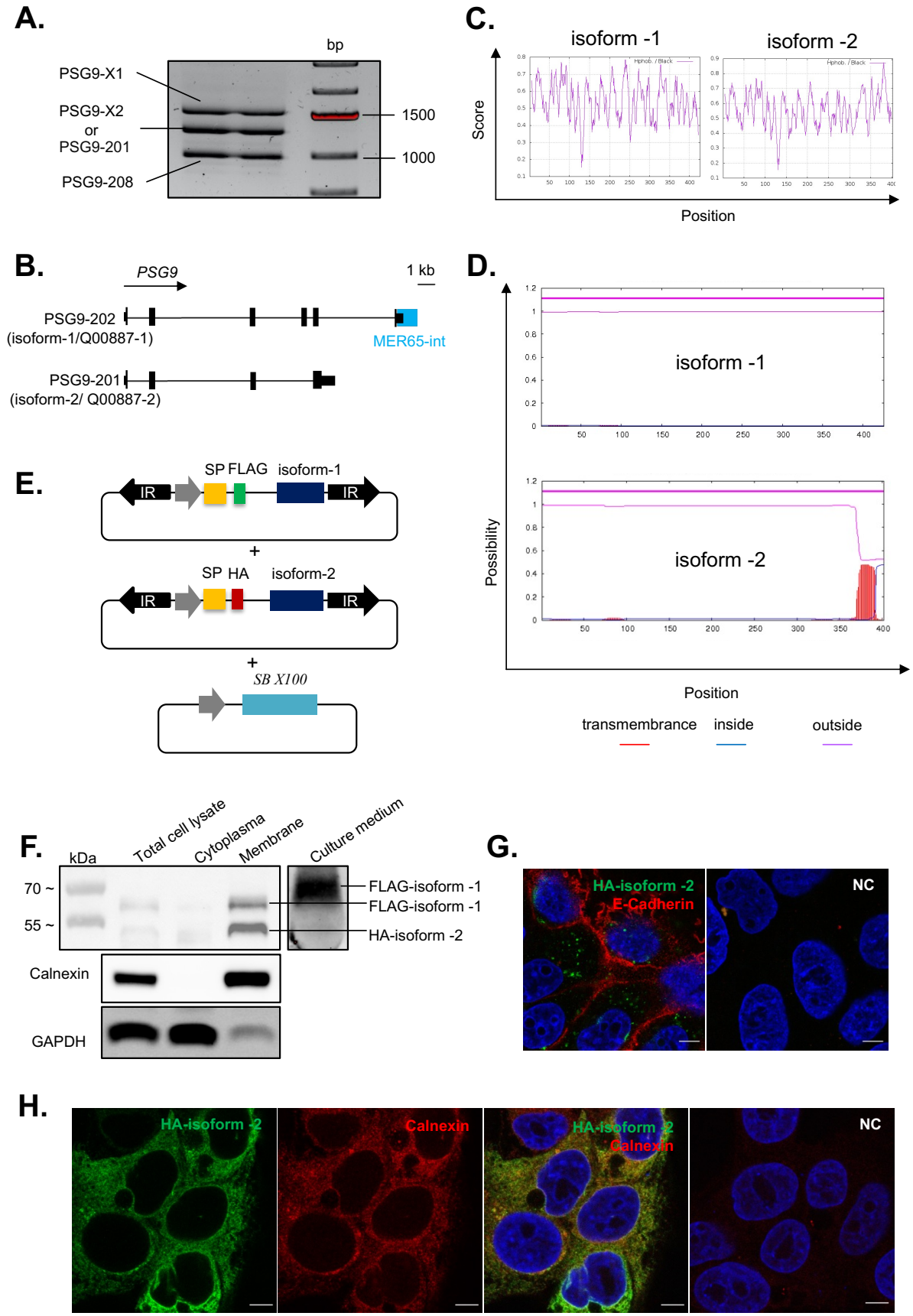

Fig. 19 The MER65-int-directed APA could lead to PSG9 isoform formation. (A) The RT-PCR analysis of subgroup C transcripts. (**B**) Diagram depicting the gene structure of two representative PSG9

isoforms. (**C**) The hydrophobicity analysis of PSG9 isoform across the amino acid position. The C-terminus of isoform -2 exhibits the higher hydrophobicity score compared to that of isoform -1. Analysis was done using the online webtool 'ProtScale' (https://web.expasy.org/protscale/) with the Hphob./Black scale [240]. (**D**) Prediction of transmembrane helices in PSG9 isoforms by the online webtool 'TMHMM Server v. 2.0' (http://www.cbs.dtu.dk/services/TMHMM/). The C-terminus of isoform -2 contains a potential transmembrane region indicated in red. (**E**) Schematic diagram of the strategy to construct the stable cell line which co-expressed the PSG9 isoforms fused to distinct tag proteins. (**F**) One representative result of the cell fractionation assay showing the FLAG-isoform -1 was detectable in cell culture media and the HA-isoform -2 was enriched in the membrane compartment. (**G**) One representative result of the immunostaining showing the expression pattern of HA-isoform -2 in trophoblasts. (**H**) One representative result of the immunostaining showing the localization of HA-isoform -1 on ER. Calnexin, an ER marker protein; GAPDH, a protein used to label the cytoplasmic compartment; E-cadherin, a protein used to label cell boundary; NC, the cells stained with the secondary antibody alone as a negative control. Scale bar, 20 μm.

3.3 The HERVs-directed PSG9 expression is required for trophoblast syncytialization

3.3.1 PSG9 expression is required for trophoblast syncytialization in *vitro*

Next, we interrogated whether the HERVs-directed PSG9 expression can exert a biological function in placentation. Given its major expression in SCTs, we first asked whether PSG9 has a role in SCT formation using the well-established BeWo syncytialization model [241,242] (Fig. 20A). Using the specific primers targeting different *PSG9* transcripts, the RT-qPCR results showed the *PSG9* mRNA was gradually increased over the BeWo cell fusion (Fig. 20B). Strikingly, the *PSG9* transcripts from subgroup B that code secretory protein constituted the dominant *PSG9* mRNA over the process (Fig. 20B). The phenomenon matched the recent report indicating the SCT differentiation elicits widespread 3'UTRs shortening of transcripts through APA process, leading to more stable mRNAs involved in secretion [243]. We further examined whether PSG9 expression is required for trophoblast syncytialization by the shRNA-directed PSG9 knockdown (PSG9 KD). As shown in Fig. 20C, the downregulation of the entire PSG9 transcripts resulted in the slight but significant decrease of canonical SCT signature genes, including *CGB*, *ERVW-1* and *ERVFRD-1*, suggesting an essential role of PSG9 for proper BeWo syncytialization. However, the comparable decrease of transcripts from either subgroup B (~ 80 %) or C (~ 50 %) alone had little

effect, meaning the potentially functional redundancy may exist between different *PSG9* transcripts (Fig. 20D).

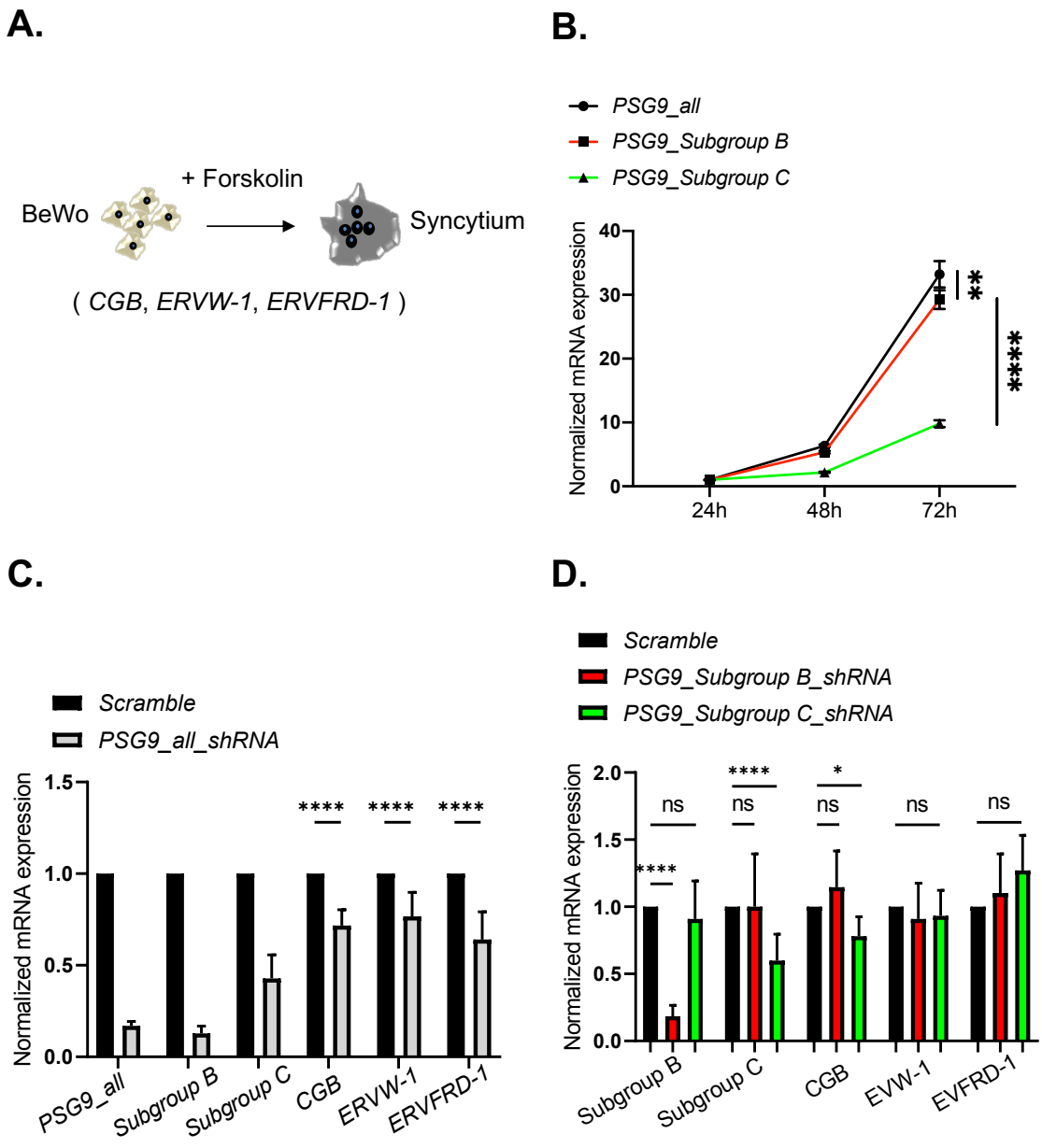

Fig. 20 PSG9 might be essential for trophoblast syncytialization. (**A**) Diagram depicting the Forskolin-induced BeWo fusion as an experimental model of trophoblast syncytialization in *vitro*. (**B**) The RT-qRT analysis of *PSG9* transcripts over BeWo syncytialization with the specific primers targeting all *PSG9* transcripts (*PSG9_all*), subgroup B (*PSG9_Subgroup B*) and C transcripts ((*PSG9_Subgroup C*), respectively. One representative result of three independent replicates. N = 3 technical replicates, mean ± SD. Unpaired t test. *PSG9*_all vs. *PSG9_Subgroup B* at 72 hr, ** $P = 0.002$; *PSG9_Subgroup B* vs. *PSG9_Subgroup C* at 72 hr, **** $P < 0.0001$. (**C**) The mRNA expression of BeWo syncytialization signature genes upon the decrease of all *PSG9* transcripts. The shRNA-directed KD effect was measured by mRNA expression of *PSG9*. N = 6 independent replicates at least, mean ±

SD. Unpaired t test. **** $P < 0.0001$. (**D**) The mRNA expression of BeWo syncytialization signature genes upon the decrease of specific *PSG9* isoform transcripts. N = 4 independent replicates at least, mean ± SD. Unpaired t test. **** $P < 0.0001$; * $P = 0.04$; ns, not significant

3.3.2 The LTR8 at *PSG9* locus is required for trophoblast syncytialization

Considering the regulatory role of the LTR8 in PSG9 expression, we further examined whether the LTR8 can exert a biological outcome in SCT formation. Indeed, the canonical signature genes of SCTs were significantly decreased in the LTR8 deletion background (LTR8 KO_P6-B4) in which *PSG9* expression was dramatically downregulated (Fig. 21A and B). By contrast, consistent with the mild phenotypes of PSG9 KD, no significant difference in gene expression was detected in the LTR8 KO colony (LTR8 KO_P5-B2) where the *PSG9* mRNA was decreased to lesser extent (Fig. 21A), suggesting again that the function of PSG9 isoform in BeWo syncytialization could be redundant. To further evaluate the fusogenic ability of BeWo cells upon LTR8 deletion, we performed the fusion assay by co-culturing the GFP- and mCherry-expressing cells and measured the percentage of resultant syncytium which displays double-positive signals by fluorescence-activated cell sorting cell sorting (FACS) (Fig. 21C). As shown in Fig. 21D, the cell size of wide type became large upon the Forskolin treatment, meaning that the cell fusion occurred and the experimental system can be used to evaluate this process. Indeed, the LTR8 KO cell population displayed smaller cell size than the wild type control (Fig. 21D). Consistently, despite the comparable ratio of the double-positive cells to the wild type in general, the percentage of cells possessing strong signals was much lower in the LTR8 KO population (LTR8 KO_P6-B4), (Fig. 21E and F), suggesting the decreased formation of large syncytium. Altogether, the genetic analysis indicated the LTR8 plays an essential role in BeWo syncytialization, including both hormone secretion and cell fusion. Nevertheless, the phenotype of LTR8 deletion cannot be rescued by introduction of either isoform alone or simultaneous overexpression of both isoform -1 and -2 (Fig. 21G), meaning the other PSG9 isoforms might be also essential for BeWo syncytialization or LTR8 contributes to the process through other alternative mechanism. Altogether, our genetic data indicated that the HERVs - regulated PSG9 expression could have a biological role in placentation through promoting trophoblast syncytialization.

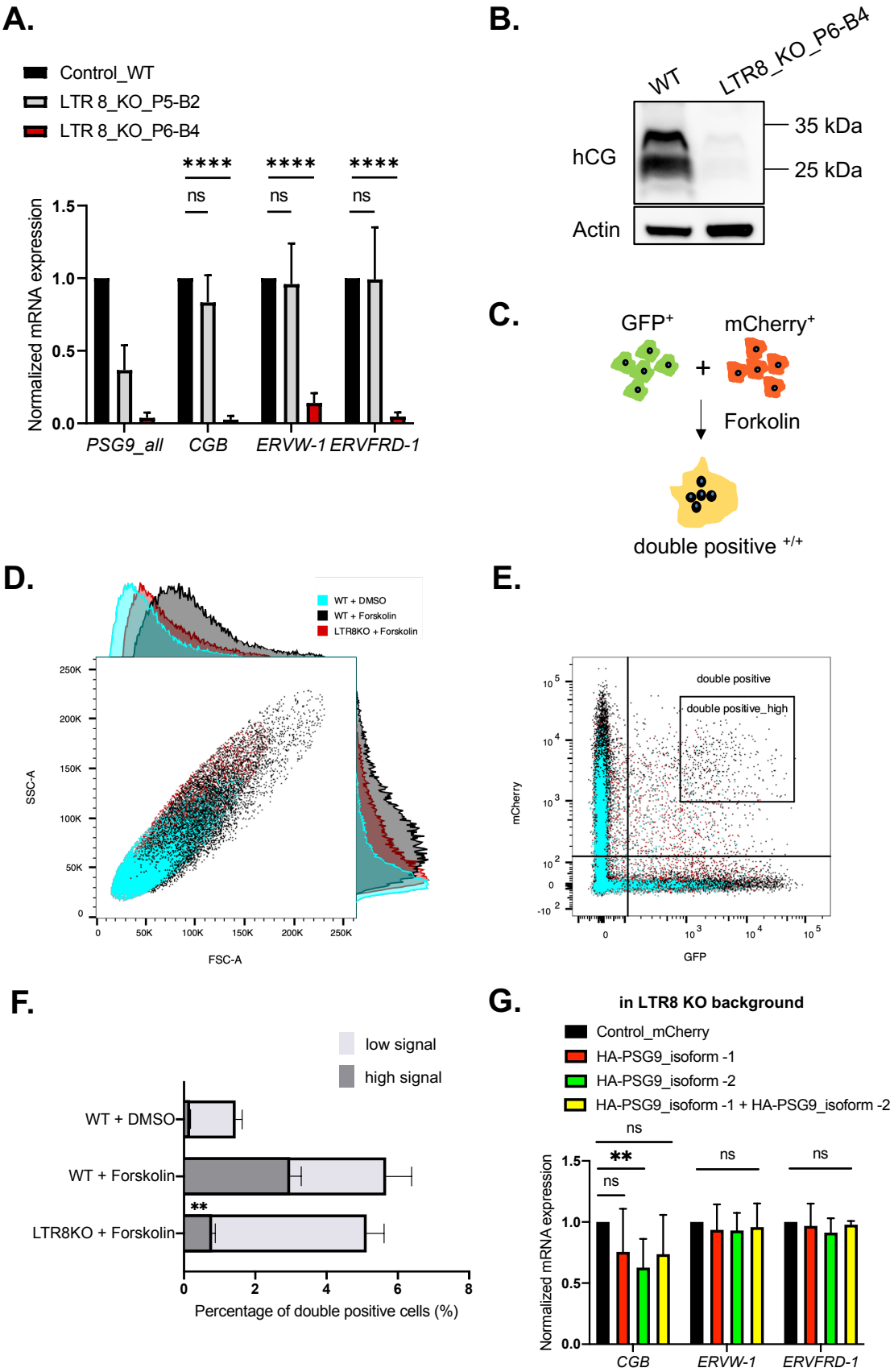

Fig. 21 The LTR8 element is essential for BeWo syncytialization. (**A**) The mRNA expression of BeWo syncytialization signature genes upon LTR8 deletion at 48 hr post-Forskolin treatment. N = 6

and 8 independent replicates for LTR8 KO colony P5-B2 and P6-B4, respectively. Mean ± SD. Unpaired t test. **** $P < 0.0001$; ns, not significant. (**B**) Western blot showing expression of hCG protein during BeWo syncytialization upon LTR8 deletion at 48 hr post-Forskolin treatment. (**C**) Schematic diagram of the strategy to measure BeWo cell fusion. The GFP- and mCherry-expressing cells were co-cultured and then treated with Forskolin. The fusion activity is measured based on the formation of syncytium containing double-positive signals (see details in methods). (**D**) One representative FACS plot showing the size of cells treated with Forskolin treatment for 48 hr. FSC-A, forward scatter area; SSC-A, side scatter area. Each dot represents a single cell. (**E**) One representative FACS plot showing the cell population displaying the double-positive fluorescence. The cells with the high signals were selected in rectangle box and represented the large syncytium. (**F**) Quantification of the syncytium displaying the double-positive fluorescence. N = 3 independent replicates, mean ± SD. Paired t test. WT high signal vs. LTR8 KO high signal, ** $P = 0.004$. (**G**) The rescue assay of LTR8 deletion by overexpression of PSG9. N = 3 independent replicates, mean ± SD. Unpaired t test. Control vs. isoform-2 for *CGB* expression, ** $P = 0.001$; control vs. each sample in the rest, not significant (ns).

3.3.3 The LTR8 at *PSG9* locus promotes the molecular signaling for trophoblast syncytialization

To gain insights into the molecular mechanisms underlying the HERVs-regulated PSG9 in SCT formation, we analyzed the transcriptome of the PSG9 KD and the LTR8 KO cells over BeWo syncytialization. We previously hypothesized that given their shared phenotypes, the commonly regulated genes in the two conditions could be more likely to participate in the signaling pathway of the HERV-directed PSG9 for syncytialization. Unexpectedly, consistent with the mild phenotypes aforementioned, only a few differentially expressed genes (DEGs) were identified between PSG9 KD and the scramble control (Fig. 22A, adjusted *p*-value < 0.01), including *RNA, U6 small nuclear 1* (*RNU6-1*), *Major vault protein* (*MVP*), *FXYD domain containing ion transport regulator 3* (*FXYD3*) and *meteorin, glial cell differentiation regulator* (*METRN*). Given the insignificant decrease of *PSG9* (log_2 fold change = -3, adjusted p value = 0.129), we speculated the mild transcriptome change might be due to the insufficient KD effect. On the other hand, in the LTR8 KO background (colony P6-B4) in which PSG9 expression was almost gone, 1976 genes and 1280 genes were identified to be significantly downregulated and upregulated, respectively (adjusted *p*-value < 0.01). Dramatically, there was comprehensive downregulation of the known SCT signature genes upon LTR8 deletion, such as *CGA*, *CGB*, *HERVW-1*, *HERVFRD-1*, *LGALS16*

and *GDF15* [243] (Fig. 22B). In addition, the mRNA levels of TSFs were decreased (Fig. 22C), including *GATA2*, *GATA3*, *TFAP2A* and *TFAP2C* [24], validating the indispensable role of the LTR8 in SCT differentiation. Nevertheless, the expression of other key players in human trophoblast stem cells, such as *CDX2* and *TEAD4*, were upregulated, suggesting the cell renewal might be increased upon the LTR8 removal (Fig. 22D).

The gene ontology (GO) analysis indicated that the downregulated genes in the LTR8 KO were dramatically enriched with the biological process associated with regulation of response to stimulus, such as positive regulation of cellular metabolic process, endoplasmic reticulum unfolded protein response (UPR), regulation of apoptotic process and cell surface receptor signaling (Table. 15). Indeed, the canonical Forskolin-stimulated cyclic AMP (cAMP) signaling pathway responsible for trophoblast syncytialization was thoroughly compromised with the downregulation of central factors (Fig. 22E), including *RAPGEF4*, *STAT5B*, *GCM1*, *OVOL1*, indicating the LTR8 at *PSG9* locus could contribute to SCT differentiation through promoting the syncytialization signaling. On the other hand, the upregulated genes seem to be mainly involved in the actin filament-based process, cell migration, cell adhesion and extracellular structure organization (Table. 16). Of note, the factors involved in trophoblast migration such as *LGALS3* and *PLAC8* were increased (Fig. 22F).

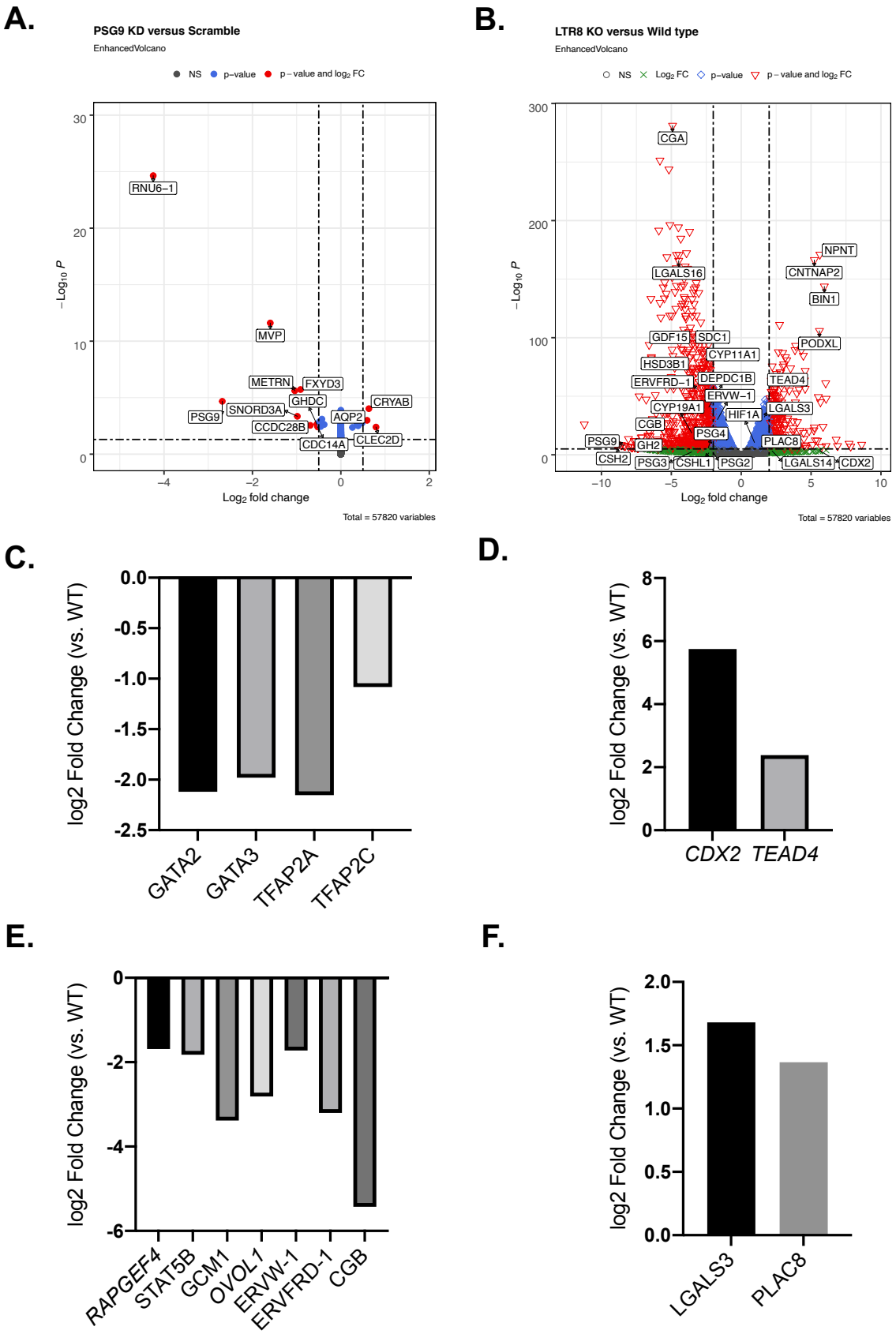

Fig. 22 The LTR8 promotes the signaling for trophoblast syncytialization. (**A**) Volcano plot visualizing the DEGs in the PSG9 KD compared to the scramble control. The X axis represents the

differential expression levels transformed into the log_2 fold change and the y axis represents the p-values attained by the Wald test without correction. The log_2 fold change cutoff = 0.5; the p value cutoff = 0.05. (**B**) Volcano plot visualizing the DEGs in the LTR8 KO compared to the wide type control. The X axis represents the differential expression levels transformed into the log_2 fold change and the y axis represents the adjusted p-values attained by the Wald test with correction for multiple testing using the Benjamini and Hochberg method. The log_2 fold change cutoff = 2; the adjusted p-value cutoff = 0.000001. (**C-F**) The bar graph showing the expression fold change of key trophoblast specification genes (**C**), key trophoblast self-renewal genes (**D**), key factors in the cAMP signaling for trophoblast syncytialization (**E**) and the candidate genes related to trophoblast migration (**F**), respectively, in the LTR8 KO compared to the wild type control.

Table. 15 The Top GO items enriched for the downregulated genes in LTR8 KO. FDR q-value, the enrichment p-value computed according to the mHG or HG model was corrected for multiple testing using the Benjamini and Hochberg method. Enrichment is defined as follows: N, the total number of genes; B, the total number of genes associated with a specific GO term; n, the number of genes in the top of the input list; b, the number of genes in the intersection; enrichment = (b/n) / (B/N) [230].

	GO Term	Description	FDR q-value	Enrichment
Biological process	GO:0007165	signal transduction	4.51E-8	1.41
	GO:0048583	regulation of response to stimulus	5.99E-7	1.38
	GO:0031325	positive regulation of cellular metabolic process	2.66E-6	1.41
	GO:0042981	regulation of apoptotic process	6.15E-6	1.65
	GO:0045766	positive regulation of angiogenesis	6.43E-6	3.30
	GO:0030968	endoplasmic reticulum unfolded protein response	2.24E-5	3.75
	GO:0007166	cell surface receptor signaling pathway	1.5E-4	1.46
Molecular function	GO:0005515	protein binding	3,00E-05	1.10
	GO:0043168	anion binding	5.4E-4	1.39
	GO:0001228	DNA-binding transcription activator activity, RNA polymerase II-specific	4.4E-3	2.07
	GO:0016301	kinase activity	2.6E-2	1.66
	GO:0005102	signaling receptor binding	7.26E-2	1.40
Cellular component	GO:0005886	plasma membrane	2.85E-3	1.29
	GO:0005737	cytoplasm	5.29E-3	1.23

Table. 16 The Top GO items enriched for the upregulated genes in LTR8 KO

Ontology	GO Term	Description	FDR q-value	Enrichment
Biological process	GO:0030029	actin filament-based process	3.06E-5	2.47
	GO:0030334	regulation of cell migration	1.78E-5	1.93
	GO:0007155	cell adhesion	7.58E-5	1.89
	GO:0043062	extracellular structure organization	5.27E-4	2.29
Molecular function	GO:0003779	actin binding	7.36E-4	2.20
	GO:0008092	cytoskeletal protein binding	5.54E-2	1.57
	GO:0042813	Wnt-activated receptor activity	5.61E-2	7.65
	GO:0005201	extracellular matrix structural constituent	5.58E-2	2.69
Cellular component	GO:0005886	plasma membrane	1.34E-9	1.47
	GO:0030054	cell junction	6.42E-8	1.95
	GO:0031012	extracellular matrix	4.07E-3	1.92
	GO:0005856	cytoskeleton	4.23E-3	1.61

3.4 Upregulated PSG9 may be involved in EO-PE pathogenesis

3.4.1 PSG9 expression is upregulated in the trophoblasts of EO-PE

The abnormal levels of human PSGs have long been reported in the blood of pregnant women with PE [129]. Significantly, PSG9 peptides were identified to be increased in the as early as 15 gestational weeks' plasma from the pregnant women who later developed EO-PE [6], suggesting the potential of PSG9 as a predictive biomarker of this disease. Despite the disturbed concentration in maternal circulation, whether PSGs expression is dysregulated in the placental cells of PE patients is to be clarified. To this end, we improved our previous transcriptomic profiling of human trophoblasts by increasing the sample size to 10 PE patients and 8 healthy controls in this study [219].

Dramatically, of ten protein-coding members in human genome, *PSG9* mRNA expression was upregulated in the trophoblasts of EO-PE (Fig. 23A). On the other hand, using the placental samples with matched gestational age between controls and patients, the RT-qPCR results confirmed *PSG9* mRNA was specially increased in EO-PE (Fig. 23B). Particularly, consistent with its dominant amount in the *PSG9* transcript pool during BeWo syncytialization, the one which encodes the secretory PSG9 isoform -1 was upregulated in EO-PE (Fig. 23C). Altogether, these data suggested the elevated levels of PSG9 in maternal circulation of EO-PE at least partially originate from its upregulation in placenta. (The transcriptome and RT-qPCR analysis were done in collaboration with Dr. Amit Pande and Dr. Florian Herse, respectively.)

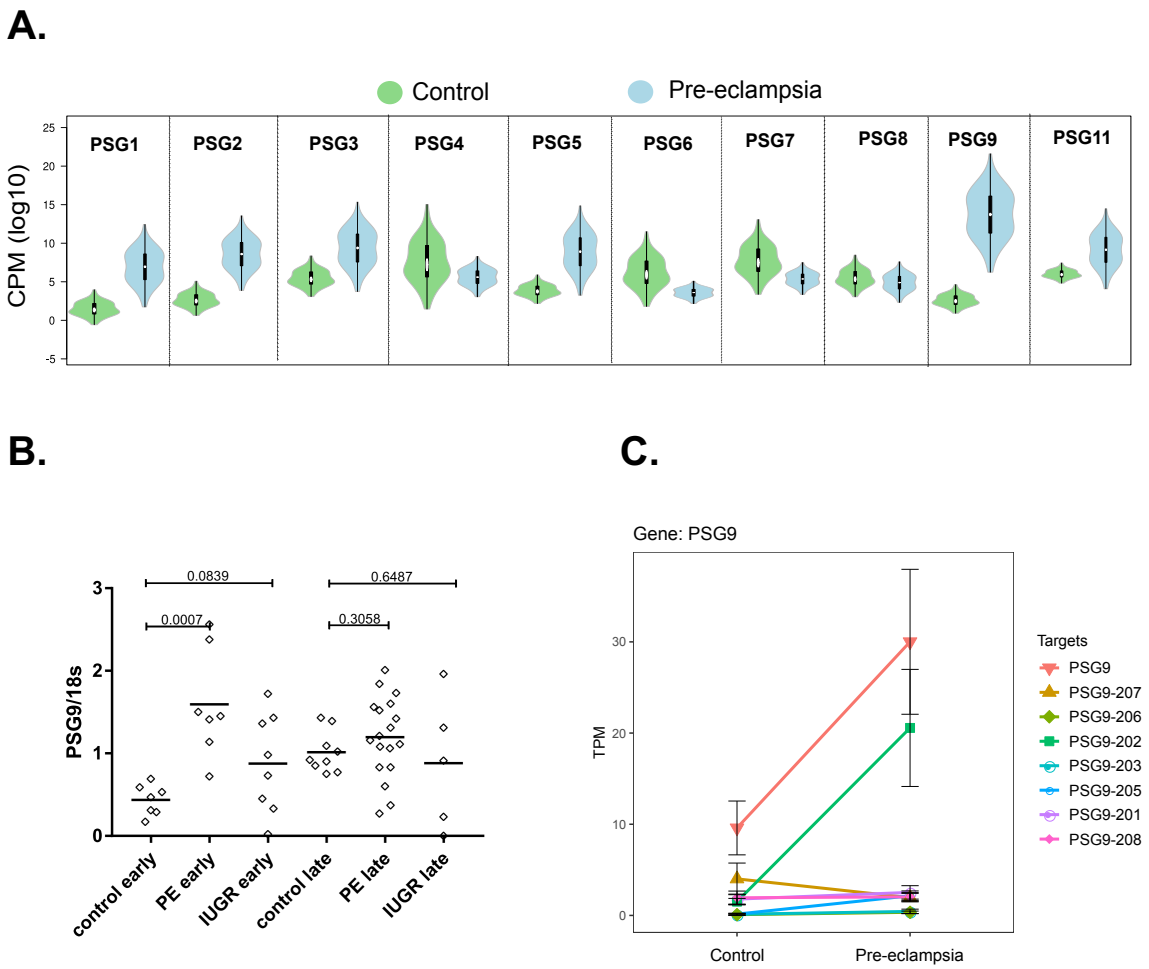

Fig. 23 *PSG9* mRNA is increased in the trophoblasts of EO-PE. (**A**) Plot showing mRNA expression of *PSG* genes in the trophoblasts of EO-PE patients and controls. CPM, counts per million. (**B**) RT-PCR analysis of *PSG9* mRNA expression in early and late onset PE, IUGR patients and healthy controls. The mRNA expression was normalized to 18s rRNA. Unpaired t test. P-value was labelled on the plot. (**C**) Expression of *PSG9* transcript isoforms in the trophoblasts of EO-PE patients and controls. The general *PSG9* transcripts is highlighted in light red and the transcript isoform *PSG9-202* which encoded the secretory PSG9 is highlighted in green. TPM, transcript per million.

3.4.2 Gain-of-function analysis of PSG9 in trophoblasts

We next interrogated whether the PSG9 increase has a role in EO-PE pathogenesis. Considering the central role of impaired placentation contributes in this disease [12,13], we performed the gain-of-function of PSG9 in human trophoblasts and examined the outcomes in trophoblast syncytialization and the invasion in *vitro*. As shown in Fig. 24A and B, the overexpression of either one isoform alone or co-expression of PSG9 isoform -1 and -2 could slightly increase the mRNA expression of CGB, meaning a

potentially positive role in promoting trophoblast syncytialization. On the other hand, the overexpression of isoform -1 had no effect on trans-well invasion of EVTs (Fig. 24C). Collectively, those data suggested that the major functions of trophoblasts may not be significantly influenced by the increase of secretory PSG9. Of interest, the ectopic isoform -2, the ER membrane-bound protein, could significantly decrease the invasive ability of EVTs in *vitro* (Fig.24C).

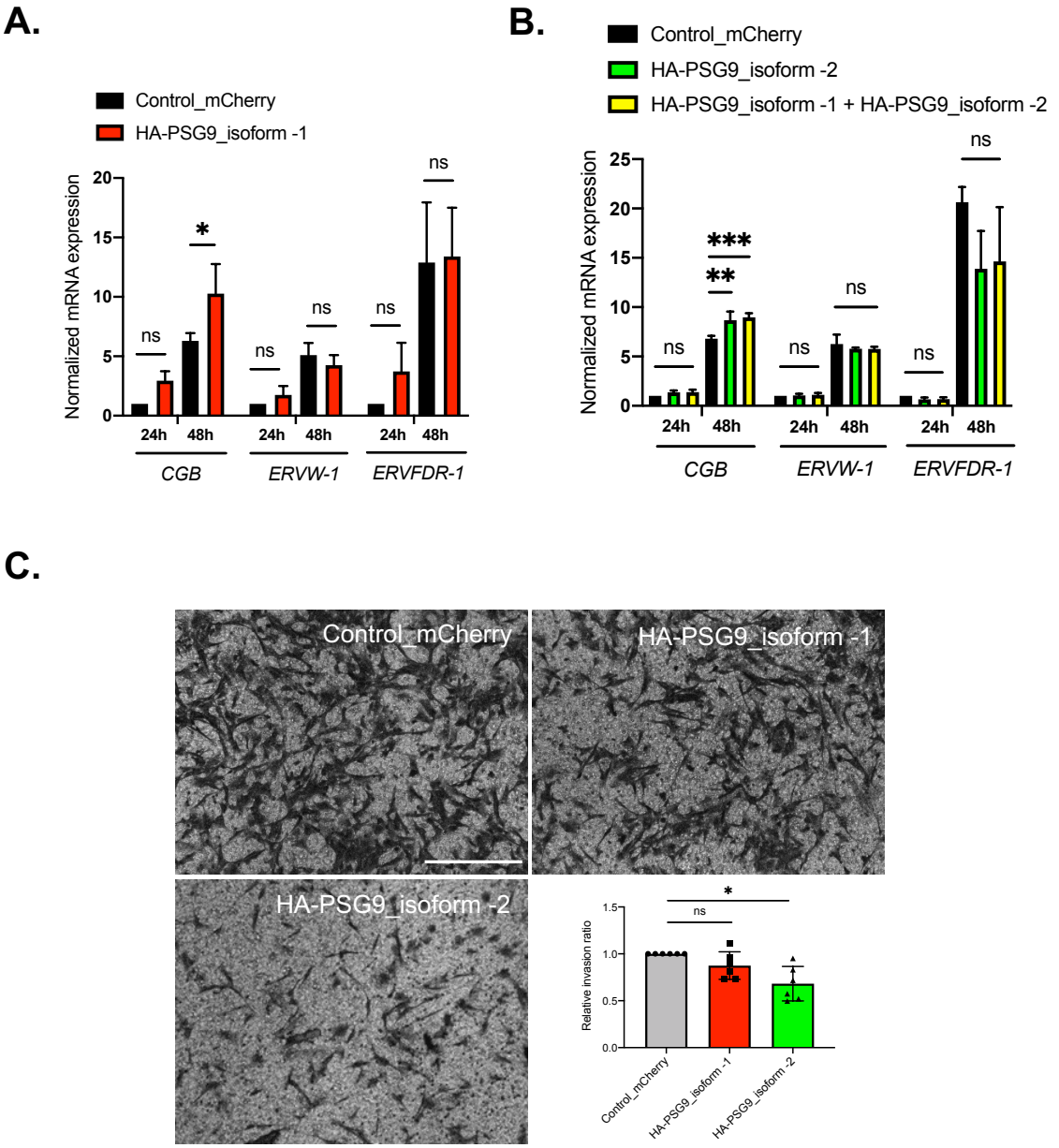

Fig. 24 Gain-of-function analysis of PSG9 in trophoblast syncytialization and invasion in *vitro*. (**A**) The RT-qPCR analysis of syncytialization marker genes upon overexpression of PSG9 isoform -1. N = 3 independent replicates, mean ± SD. Two-way ANOVA followed by multiple comparations.

Control vs. isoform -1 at 48 hr, * $P = 0.029$; the rest, not significant (ns). (**B**) The RT-qPCR analysis of syncytialization marker genes upon either overexpression of PSG9 isoform -2 or co-expression of PSG9 isoform -1 and -2. N = 3 independent replicates, mean ± SD. Two-way ANOVA followed by multiple comparations. Control vs. isoform -2 at 48 hr, ** $P = 0.002$; control vs. co-expression, *** $P = 0.0005$; the rest, not significant (ns). (**C**) One presentative result of PSG9 overexpression in the trans-well invasion with EVTs in *vitro.* N = 6 independent replicates with 4 technical counting from non-overlapping microscopic views each, mean ± SD. RM one-way ANOVA with followed by multiple comparations. Control vs. isoform -1, not significant; control vs. isoform -2, * $P = 0.014$. Scale bar, 200 μm.

4. Discussion

PSG proteins have been recognized in maternal blood during pregnancy since 1970s [2,114]. Despite the implications in placental physiology and related diseases, the regulatory mechanism and biological function of PSG expression in human trophoblast have been poorly elucidated. With the lake of specific antibody targeting a given member, to characterize expression patterns of different PSG proteins in placenta has been further hindered. Thanks to comprehensive genome annotation and advanced sequencing technologies, our understanding of HERVs as regulatory elements in placenta has been largely improved in the past decade, allowing us to investigate PSGs from a novel perspective. In this study, we systematically examined regulatory roles of HERVs in PSG expression and identified two elements, LTR8 and MER65-int, which could control PSG9 expression at different levels (Fig. 25). Importantly, the HERVs-regulated gene expression could exert a biological role in placentation through facilitating trophoblast syncytialization. On the other hand, this study revealed that the secretory PSG9 is upregulated in trophoblasts of EO-PE and supported the notion that PSG9 could be used as a potential biomarker to predict this disease. Given the prevalent distribution of the two elements across the entire *PSG* gene cluster, the mechanisms underlying the regulation of HERVs in PSG9 expression bring insights into future research of other PSG members.

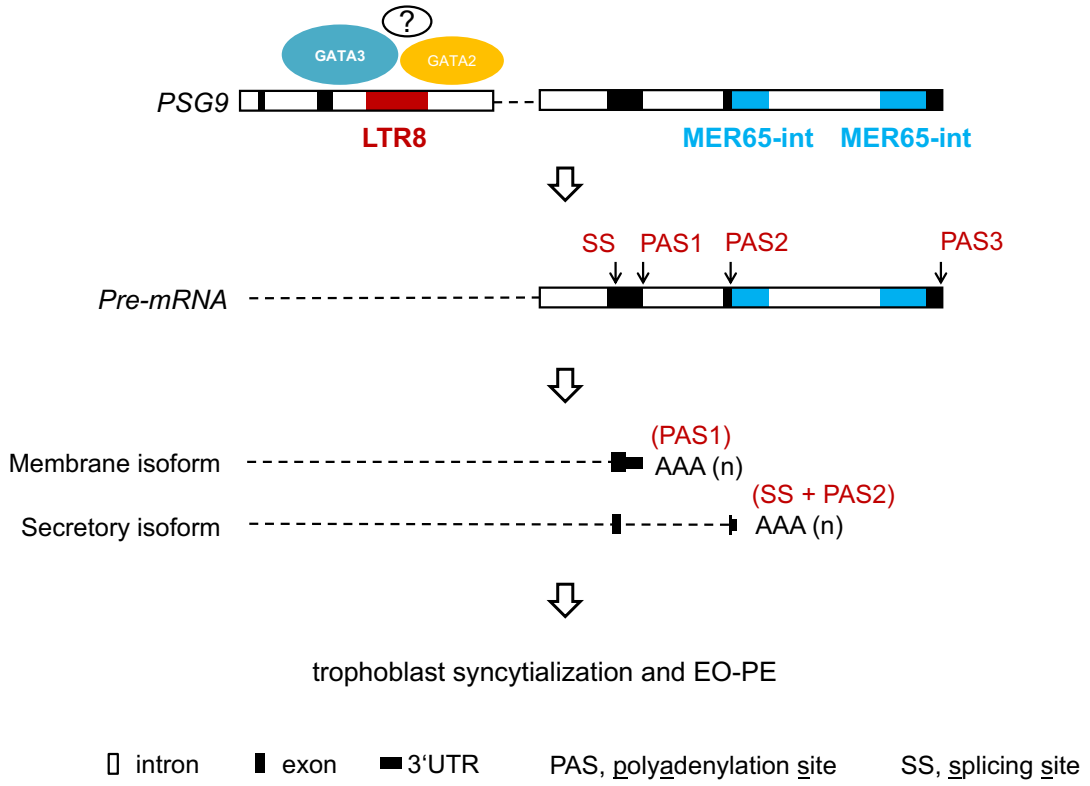

Fig. 25 The proposed working model of HERV elements-directed PSG9 in placentation and PE. The LTR8 and MER65-int could regulate PSG9 expression at different levels. The LTR8 promotes *PSG9* transcription as an enhancer bound by GATA2, GATA3 or other unknown factors. On the other hand, working together with the alternative splicing process, the MER65-int could contribute to alternative polyadenylation of *PSG9* mRNA by providing *cis*-elements and/or pA sites, leading to PSG9 isoforms with different localizations. The HERVs-directed PSG9 could play a biological role in placentation through promoting trophoblast syncytialization, while the dysregulation of PSG9 may be implicated in EO-PE pathogenesis.

4.1 HERVs-derived elements control PSG9 expression at both transcriptional and post-transcriptional levels

Accumulating evidence has revealed that HERVs-derived sequences are frequently co-opted in placenta, serving as an abundant source of transcriptional modulatory elements [8,184,211-213]. Here we presented an elegant case showing HERVs can modulate the same gene expression in a versatile manner. The LTR8 promotes PSG9 transcription as an enhancer bound by GATA2 and GATA3, whereas the MER65-int

could contribute to the APA of *PSG9* mRNA at post-transcriptional level, producing PSG9 isoforms with different localizations (Fig. 25).

4.1.1 The LTR8 promotes PSG9 transcription as a placenta-specific enhancer bound by trophoblast transcription factors GATA2 and GATA3

Mining the online available epigenetic data, we found the LTR8 copies embedded into human *PSG* gene clusters are enriched with the enhancer-associated chromatin features in trophoblasts. Indeed, the LTR8 at *PSG9* locus is able to exhibit robust enhancer activity in the luciferase reporter assay. Importantly, PSG9 mRNA is downregulated upon the CRISPR-directed LTR8 deletion, suggesting that the HERV-derived LTR8 element is required for endogenous PSG9 expression. The LTR8 controls PSG9 transcription as an enhancer with binding to the TSFs, including GATA2 and GATA3. However, the overexpression of GATA3 can increase *PSG9* mRNA independent of LTR8 occurrence, suggesting the LTR8 might be sufficient not necessary for ectopic GATA3-directed PSG9 expression. Nevertheless, given the exogenous nature of GATA3 overexpression, we cannot conclude here whether the LTR8 is required for the endogenous GATA3-directed PSG9 expression. The CRISPR-directed specific manipulation of the binding site of GATA3 on the LTR8 element holds promise to address it. Interestingly, PSG1 expression cannot respond to ectopic TSFs, despite the analogous chromatin landscape of LTR8 across the entire *PSG* family, meaning the expression of different PSG member could be divergently regulated. Overall, whether the modulation model ‘LTR8-TSFs’ contributes to transcription of other PSG needs to be examined experimentally case by case.

The LTR8 at *PSG9* locus is especially present in the genome of anthropoid primates. Of note, the chromatin of LTR8 copies in human trophoblasts displays more accessibility compared with that in other cell types, suggesting the LTR8 could be a placenta-specific enhancer in simians. In addition, the online available Hi-C data using human trophoblast lineage indicate PSGs and the closely related CEACAMs may belong to the same TAD, where the LTR8 potentially interacts with other genomic regions, suggesting a potentially broad role of the LTR8 in gene regulation other than controlling PSG9 expression. However, a clear interactome of the LTR8 element across human genome during trophoblast differentiation needs to be deciphered in the future.

4.1.2 The MRE65-int contributes to alternative polyadenylation of *PSG9* mRNA and isoform formation

Despite the considerable divergence of PSG variants, we found *PSG9* transcripts can be divided into three groups depending on whether the 3'UTR region merges with the MER65-int or not. The MER65-int may contribute to mRNA polyadenylation of PSG9 by providing PASs or *cis*-elements. Indeed, by performing the polyadenylation reporter assay, we found the polyadenylation efficiency of target mRNA is significantly compromised upon the removal of the full-length MER65-int, indicating the MER65-int is essential for sufficient mRNA polyadenylation. The site-specific mutagenesis, such as CRISPR-based gene editing [244,245], would provide further evidence to confirm the endogenous role of MER65-int in polyadenylation of *PSG9* mRNA.

Besides the occurrence at *PSG9* locus, the MER65-int element is prevalent across the entire human *PSG* gene cluster and overlaps with the 3'UTR regions of all other *PSG* members. Indeed, compared with the counterpart alignments in the MER65-int consensus sequence [246], the copies in the *PSG* gene cluster are more likely to contain the core upstream *cis*-element (AAUAAA) essential for mRNA polyadenylation. Interestingly, it is worth noting that in addition to the insertion at 3' terminus, we found the MER65-int is present in the upstream region of several *PSG* genes, including *PSG8, PSG7, PSG11, PSG2, PSG4* and *PSG9* (GRCh38/hg38). Given its specific distribution pattern, it is tempting to speculate that the capture of MER65-int in human genome might have a role in shaping *PSG* gene structure over the course of evolution.

Working together with the alternative splicing process, the MER65-int seems to particularly contribute to the APA of *PSG9* mRNA. It is well known that the APA-directed transcript isoforms could vary in mRNA metabolism such as stability, localization and translation rate [237,238]. Here we found the MER65-int-directed APA could contribute to PSG9 isoforms with different localizations. The PSG9 isoforms with the C-terminuses merging the MER65-int are secretory, whereas the ones containing long hydrophobic C-terminuses where the intronic polyadenylation occurs are intracellular and retained at the ER membrane. Given that PSGs have long been recognized to be secreted proteins, it would be interesting to interrogate whether the ER membrane-bound isoform is biologically functional. Indeed, the overexpression of the ER membrane-bound PSG9 can result in significantly decreased invasion of EVTs

in *vitro*, although the underlying mechanism needs to be further investigated. Furthermore, to interrogate functional interaction between different PSG9 isoforms during placentation is of great interest. This idea was inspired by the case of imbalance between sFLT1 and its splicing variant FLT1 in PE, in which the excess soluble sFLT1 antagonizes the cell membrane-bound FLT1 to bind to the circulating proangiogenic factors VEGF and PLGF, leading to endothelial cell dysfunction [97-100]. Besides, the functional divergence of two main splice forms of CEACAM1, a paralogous member of PSGs, has been well characterized. The isoforms with the long cytoplasmic tail are mainly expressed in immune cells and provide inhibitory signals via the immunoreceptor tyrosine-based inhibitory motifs (ITIMs) and immunoreceptor tyrosine-based switch motifs (ITSMs), while the short isoforms lack of the ITIM/ITSM motifs are expressed in epithelial cells [111,247,248].

In addition, our experimental results revealed that the transcripts that encode secretory PSG9 are overrepresented during BeWo syncytialization. One important question next is to identify what factor determining the expression ratio of PSG9 isoforms over this process. It has been proposed that specific APA events could occur in a condition or cell type-dependent manner [237]. The condition-dependent expression of regulators that bind to sequence motifs or *cis*-elements in pre-mRNAs can influence which PAS is used. For instance, the increase of CstF-64, a general polyadenylation processing factor [249], enhances an effective utilization of the proximal PAS in a composite terminal exon, leading to a switch from the membrane-bound to the secreted form of immunoglobulin M heavy chain during B cell activation [250]. Another possibility is the selective degradation of *PSG9* transcripts. Consistently, a recent report revealed that the intronic polyadenylation isoforms are generally less stable than isoforms whose PASs are in the 3' terminal exon over SCT differentiation [243]. In both scenarios, the genetic screening of regulators which interact with the *PSG9* pre-mRNA, such as microRNAs and RNA-binding proteins (RBPs) [237], would provide more insights into the dynamic turnover of *PSG9* transcripts.

4.2 The HERVs-directed PSG9 expression is required for trophoblast syncytialization

Despite the involvements in multiple physiological processes in pregnancy, the biological function of PSGs in trophoblast differentiation has been rarely explored. In

this study, our data suggested the HERVs-regulated PSG9 could contribute to normal placentation by promoting SCT formation.

Firstly, the mRNA levels of PSG9 are gradually increased over BeWo syncytialization, where the transcripts encoding the secretory PSG9 are predominantly expressed. As discussed above, to investigate the regulatory mechanism controlling the expression pattern of PSG9 isoforms is an interesting direction. Next, the general downregulation of PSG9 could slightly inhibit BeWo syncytialization based on the decreased expression of SCT signature genes, whereas the specific knock-down (KD) of a defined isoform has no effect. In accordance with this, the transcriptome of PSG9 KD is nearly indistinguishable from that of the scramble control. The mild phenotype of PSG9 KD likely results from functional redundancy of PSG9 isoform and/or insufficient KD effect. A more robust method, such as the Cas13 proteins-targeted RNA KD [251], could be applied to validate these findings.

On the other hand, in line with the positive regulation of the LTR8 in PSG9 expression in general, the loss-of-function of LTR8 causes defective BeWo syncytialization. Dramatically, the phenotype is conspicuous in the LTR8 KO background where PSG9 expression is virtually absent, suggesting an essential role of the LTR8-directed PSG9 expression in SCT differentiation. Nevertheless, overexpression of the representative PSG9 isoforms fails to rescue the phenotype of LTR8 deletion, which means other isoforms might also be required or the LTR8 contributes to trophoblast syncytialization by affecting alternative pathways other than modulating PSG9 expression. The transcriptomic profiling further showed that the classical cAMP signalling for trophoblast syncytialization is comprehensively compromised upon LTR8 deletion, whereas the self-renewal genes of trophoblast stem cells, including *CDX2* and *TEDA4*, are upregulated. Taken together, these data prompted us to propose that the HERV elements contribute to trophoblast differentiation, at least partially, through orchestrating the expression pattern of PSG9 isoforms, whereby the highly expressed soluble PSG9 likely acts as a ligand to enhance the cAMP signalling for SCT specification. This working model mirrors the ‘positive-feedback loop’ of hCG during SCT differentiation, in which hCG synthesized in SCTs binds to the receptor LHCGR and further reinforce the cAMP/PKA signalling [10,44,45].

Next, one critical question would be to identify the cell-surface receptor or interactor of PSG9 involving in the signalling pathway for syncytialization. Human PSGs, such

as PSG1, have been revealed to bind to heparan sulfate proteoglycans, integrins and Gal-1, influencing multiple physiological processes in placentation including angiogenesis, trophoblast migration and immune response [3-5,126]. Notably, the soluble Gal-1 has been reported to promote trophoblast syncytialization through stabilizing binding of Syncytin-2 to its receptor MFSD2a [252]. Thus, it is tempting to interrogate whether PSG9 also acts as a ligand interacting with Gal-1 in this process (discussion with Prof. Dr. Gabriela S. Dveksler).

Despite the promising results, one limitation of this study is that the BeWo syncytialization system alone was used. Efficient PSG9 KD and LTR8 KO in the recently developed hTSCs [67,71-73] or even the 3D human trophoblast organoids [74,75] would bring more insights into the molecular mechanisms underpinning trophoblast differentiation. More importantly, these advanced tools would provide unprecedented opportunities to examine the regulatory role of the HERVs-directed PSG9 expression in human primary SCT specification during the peri-implantation stage. With the lack of a suitable study system, the trophoblast differentiation at that stage has puzzled researchers [10 253]. On the other hand, it is also intriguing to investigate whether other PSG members play a similarly biological role to that of PSG9 in placentation.

4.3 PSG9 is upregulated in the placenta of EO-PE

The abnormal concentration of PSG proteins in maternal blood has long been implicated in pregnancy complications including PE [129]. Lacking a specific antibody, our knowledge about the expression pattern of a given PSG member over PE development has been rather limited. Using the mass spectrometry related techniques, a recent report identified a subset of PSGs, dramatically PSG9, are increased in the plasma samples of EO-PE as early as the 15th week before clinical symptoms emerge [6,254], proposing PSG9 as a potentially predictive biomarker of EO-PE. Despite the promising candidate, it remains unclear that whether the elevated PSG9 originates from its dysregulation in trophoblasts or the release by more apoptotic placental cells shed into maternal circulation [80,255-258]. With the transcriptomic profiling and the RT-qPCR confirmation, our data revealed that the secretory PSG9 isoform is particularly upregulated in the trophoblasts of EO-PE. Consistently, the previous report from our lab has shown GATA3, a master regulator of PSG9 transcription identified in this study, is overexpressed in EO-PE [219]. Given the PSG9 increase induced by GATA3

overexpression, we proposed that the hyperactive 'GATA3-LTR8' module might contribute to PSG9 upregulation in the trophoblasts of EO-PE. Next, an important question is to identify what trophoblast subtype expresses higher levels of PSG9 in EO-PE. To decipher the single-cell transcriptome of EO-PE placenta together with immunostaining would shed light on this direction. On the other hand, due to the limited availability of placental samples in the first and second trimester, this study cannot determine when the PSG9 dysregulation occurs over EO-PE progression. With the maternal blood across different gestational age, the circulating RNA profiling and specific ELISA assay could be informative [259]. Taken together, to decipher the spatio-temporal expression pattern of PSG9 would provide insights into its role in this disease.

It is worth noting that PSG9 dysregulation in EO-PE might also be associated with other genetic or epigenetic factors. Indeed, a copy number deletion of *PSG11* (*PSG9*, based on GRCh37/hg19 or GRCh38/hg38) in maternal genome has been identified to be enriched in PE cases (5/169) and compared with controls (1/114) [260]. Besides, the implication of epigenetic disturbance, particularly DNA methylation, in PE has been extensively documented [261-264]. However, the roles of these factors in PSG9 dysregulation need to be further clarified.

Interestingly, in conflict with the increase of PSG9, the concentration of PSG1 has been identified to be lower in the maternal blood of both EO-PE and LO-PE [5], suggesting again different PSG members might vary in expression regulation and biological function in normal placentation and disease.

4.4 Increased PSG9 is associated with EO-PE

It has been widely recognized that EO-PE could originate from poor placentation even in the first trimester, particularly deficient remodelling of the maternal uterine spiral arteries by the invasive EVTs [12,13]. To interrogate the pathological role of PSG9 upregulation in EO-PE, we examined the phenotype of PSG9 overexpression in both trophoblast syncytialization and the invasion in *vitro*. Despite the mild role in promoting BeWo syncytialization, overexpression of secretory PSG9 has no effect in the invasive ability of EVTs in *vitro.* By contrast, a recent report showed the treatment with recombinant PSG9 could decrease trophoblast migration in a 3D motility model [152].

Despite a potentially causal role in EO-PE pathogenesis, we cannot totally renounce the argument that PSG9 upregulation is not a cause but a consequence of this disease. Indeed, PSGs have been proposed to contribute to several important processes in placentation for a successful pregnancy [3-5,126]. Consistently, our experimental data suggested that the HERVs-directed PSG9 could be required for trophoblast syncytialization. It is likely that the PSG9 upregulation is one of compensatory effect that the placenta exerts to combat with the physiological disturbance caused by EO-PE. In accordance with this, the enhanced syncytialization and macropinocytosis have been recently detected in the placenta with fetal growth restriction (FGR), which has been recognized as an adaptive response to sustain fetal development under nutrient scarcity [82].

To gain a better understanding of the role of PSG9 dysregulation in EO-PE, a robust system to evaluate the outcomes of its loss or gain-of-function in placentation is needed. Efforts are ongoing to develop the disease models of EO-PE, including the cell- and animal-based ones. Although no single model can recapitulate all aspects of EO-PE, a human or certain primates-specific syndrome, they are valuable tools to study specific aspects of this disease [86]. For instance, the animal models, particularly mouse and rat models, have been applied to examine various biomarkers of EO-PE, including sFlt-1 [101-103]. Treating the pregnant animals with recombinant PSG9 could offer insights into its pathological role. On the other hand, the gain-of-function of PSG9 in the hTSCs or trophoblast organoids which are developed from EO-PE pregnancy could also be a promising method to address these puzzles [265].

5. Conclusions

This study revealed that the HERV-derived sequences could regulate PSG9 expression at both transcriptional and post-transcriptional levels, providing insights into the versatile roles of HERVs in gene regulation. Importantly, the HERVs-directed PSG9 expression could exert a biological function in trophoblast syncytialization, supporting the notion that the co-option of HERVs contributes to placental development and evolution. On the other hand, the expression of secretory PSG9 isoform is upregulated in the trophoblasts of EO-PE. The better understanding of regulation and function behind the HERVs-directed PSG9 expression in placenta holds promise to develop the novel biomarker and treatment method for EO-PE.

6. References

1 Moore, T. & Dveksler, G. S. Pregnancy-specific glycoproteins: complex gene families regulating maternal-fetal interactions. *Int J Dev Biol* **58**, 273-280, doi:10.1387/ijdb.130329gd (2014).

2 Nurbubu T. Moldogazieva, I. M. M. a. A. A. T. Pregnancy-Specific β1-Glycoproteins: Combined Biomarker Roles, Structure/Function Relationships and Implications for Drug Design. *Current Medicinal Chemistry* **23**, 1-23, doi:10.2174/0929867324666161123 (2016).

3 Rattila, S., Kleefeldt, F., Ballesteros, A., Beltrame, J. S., M, L. R., Ergün, S. & Dveksler, G. Pro-angiogenic effects of pregnancy-specific glycoproteins in endothelial and extravillous trophoblast cells. *Reproduction* **160**, 737-750, doi:10.1530/rep-20-0169 (2020).

4 Warren, J., Im, M., Ballesteros, A., Ha, C., Moore, T., Lambert, F., Lucas, S., Hinz, B. & Dveksler, G. Activation of latent transforming growth factor-β1, a conserved function for pregnancy-specific beta 1-glycoproteins. *MHR: Basic science of reproductive medicine* **24**, 602-612, doi:10.1093/molehr/gay044 (2018).

5 Rattila, S., Dunk, C. E. E., Im, M., Grichenko, O., Zhou, Y., Yanez-Mo, M., Blois, S. M., Yamada, K. M., Erez, O., Gomez-Lopez, N., Lye, S. J., Hinz, B., Romero, R., Cohen, M. & Dveksler, G. Interaction of Pregnancy-Specific Glycoprotein 1 With Integrin Alpha5beta1 Is a Modulator of Extravillous Trophoblast Functions. *Cells* **8**, doi:10.3390/cells8111369 (2019).

6 Blankley, R. T., Fisher, C., Westwood, M., North, R., Baker, P. N., Walker, M. J., Williamson, A., Whetton, A. D., Lin, W., McCowan, L., Roberts, C. T., Cooper, G. J. S., Unwin, R. D. & Myers, J. E. A Label-free Selected Reaction Monitoring Workflow Identifies a Subset of Pregnancy Specific Glycoproteins as Potential Predictive Markers of Early-onset Pre-eclampsia. *Molecular & Cellular Proteomics* **12**, 3148-3159, doi:10.1074/mcp.m112.026872 (2013).

7 Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C. & Heidmann, T. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120507, doi:10.1098/rstb.2012.0507 (2013).

8 Chuong, E. B. The placenta goes viral: Retroviruses control gene expression in pregnancy. *PLOS Biology* **16**, e3000028, doi:10.1371/journal.pbio.3000028 (2018).

9 Hemberger, M., Hanna, C. W. & Dean, W. Mechanisms of early placental development in mouse and humans. *Nature Reviews Genetics*, doi:10.1038/s41576-019-0169-4 (2019).

10 Turco, M. Y. & Moffett, A. Development of the human placenta. *Development* **146**, dev163428, doi:10.1242/dev.163428 (2019).

11 Xiao, Z., Yan, L., Liang, X. & Wang, H. Progress in deciphering trophoblast cell differentiation during human placentation. *Curr Opin Cell Biol* **67**, 86-91, doi:10.1016/j.ceb.2020.08.010 (2020).

12 Phipps, E. A., Thadhani, R., Benzing, T. & Karumanchi, S. A. Pre-eclampsia: pathogenesis, novel diagnostics and therapies. *Nature Reviews Nephrology* **15**, 275-289, doi:10.1038/s41581-019-0119-6 (2019).

13 Burton, G. J., Redman, C. W., Roberts, J. M. & Moffett, A. Pre-eclampsia: pathophysiology and clinical implications. *BMJ*, l2381, doi:10.1136/bmj.l2381 (2019).

14 Schlafke, S. & Enders, A. C. Cellular Basis of Interaction Between Trophoblast and Uterus at Implantation. *Biology of Reproduction* **12**, 41-65, doi:10.1095/biolreprod12.1.41 (1975).

15 Knöfler, M., Haider, S., Saleh, L., Pollheimer, J., Gamage, T. K. J. B. & James, J. Human placenta and trophoblast development: key molecular mechanisms and model systems. *Cellular and Molecular Life Sciences* **76**, 3479-3496, doi:10.1007/s00018-019-03104-6 (2019).

16 Graham J.Burton, E., Adrian L.Watson. Maternal arterial connections to the placental intervillous space during the first trimester of human pregnancy: the Boyd collection revisited. *Am J Obstet Gynecol* **181** (1999).
17 Watts, J. L. & Ralston, A. Universal assembly instructions for the placenta. *Nature*, doi:10.1038/d41586-020-02914-0 (2020).
18 Kusama, K., Bai, R. & Imakawa, K. Regulation of human trophoblast cell syncytialization by transcription factors STAT5B and NR4A3. *Journal of Cellular Biochemistry* **119**, 4918-4927, doi:10.1002/jcb.26721 (2018).
19 Gerri, C., McCarthy, A., Alanis-Lobato, G., Demtschenko, A., Bruneau, A., Loubersac, S., Fogarty, N. M. E., Hampshire, D., Elder, K., Snell, P., Christie, L., David, L., Van de Velde, H., Fouladi-Nashta, A. A. & Niakan, K. K. Initiation of a conserved trophectoderm program in human, cow and mouse embryos. *Nature*, doi:10.1038/s41586-020-2759-x (2020).
20 Nishioka, N., Inoue, K.-I., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R. O., Ogonuki, N., Makita, R., Kurihara, H., Morin-Kensicki, E. M., Nojima, H., Rossant, J., Nakao, K., Niwa, H. & Sasaki, H. The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass. *Developmental Cell* **16**, 398-410, doi:10.1016/j.devcel.2009.02.003 (2009).
21 Niakan, K. K. & Eggan, K. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Developmental Biology* **375**, 54-64, doi:10.1016/j.ydbio.2012.12.008 (2013).
22 Berg, D. K., Smith, C. S., Pearton, D. J., Wells, D. N., Broadhurst, R., Donnison, M. & Pfeffer, P. L. Trophectoderm Lineage Determination in Cattle. *Developmental Cell* **20**, 244-255, doi:10.1016/j.devcel.2011.01.003 (2011).
23 Ralston, A., Cox, B. J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J. S. & Rossant, J. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* **137**, 395-403, doi:10.1242/dev.038828 (2010).
24 Krendl, C., Shaposhnikov, D., Rishko, V., Ori, C., Ziegenhain, C., Sass, S., Simon, L., Müller, N. S., Straub, T., Brooks, K. E., Chavez, S. L., Enard, W., Theis, F. J. & Drukker, M. 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**, E9579-E9588, doi:10.1073/pnas.1708341114 (2017).
25 Kuckenberg, P., Kubaczka, C. & Schorle, H. The role of transcription factor Tcfap2c/TFAP2C in trophectoderm development. *Reproductive BioMedicine Online* **25**, 12-20, doi:10.1016/j.rbmo.2012.02.015 (2012).
26 Soncin, F., Khater, M., To, C., Pizzo, D., Farah, O., Wakeland, A., Arul Nambi Rajan, K., Nelson, K. K., Chang, C.-W., Moretto-Zita, M., Natale, D. R., Laurent, L. C. & Parast, M. M. Comparative analysis of mouse and human placentae across gestation reveals species-specific regulators of placental development. *Development* **145**, dev156273, doi:10.1242/dev.156273 (2018).
27 Adachi, K., Nikaido, I., Ohta, H., Ohtsuka, S., Ura, H., Kadota, M., Wakayama, T., Hiroki & Niwa, H. Context-Dependent Wiring of Sox2 Regulatory Networks for Self-Renewal of Embryonic and Trophoblast Stem Cells. *Molecular Cell* **52**, 380-392, doi:10.1016/j.molcel.2013.09.002 (2013).
28 Latos, P. A., Goncalves, A., Oxley, D., Mohammed, H., Turro, E. & Hemberger, M. Fgf and Esrrb integrate epigenetic and transcriptional networks that regulate self-renewal of trophoblast stem cells. *Nature Communications* **6**, 7776, doi:10.1038/ncomms8776 (2015).
29 Kunath, T., Yamanaka, Y., Detmar, J., MacPhee, D., Caniggia, I., Rossant, J. & Jurisicova, A. Developmental differences in the expression of FGF receptors between human and mouse embryos. *Placenta* **35**, 1079-1088, doi:10.1016/j.placenta.2014.09.008 (2014).
30 Horii, M., Li, Y., Wakeland, A. K., Pizzo, D. P., Nelson, K. K., Sabatini, K., Laurent, L. C., Liu, Y. & Parast, M. M. Human pluripotent stem cells as a model of trophoblast differentiation in both normal development and disease. *Proceedings of the National Academy of Sciences* **113**, E3882-E3891, doi:10.1073/pnas.1604747113 (2016).
31 Hemberger, M., Udayashankar, R., Tesar, P., Moore, H. & Burton, G. J. ELF5-enforced transcriptional networks define an epigenetically regulated trophoblast stem cell compartment

in the human placenta. *Human Molecular Genetics* **19**, 2456-2467, doi:10.1093/hmg/ddq128 (2010).

32 Lee, Y., Kim, K. R., McKeon, F., Yang, A., Boyd, T. K., Crum, C. P. & Parast, M. M. A unifying concept of trophoblastic differentiation and malignancy defined by biomarker expression. *Hum Pathol* **38**, 1003-1013, doi:10.1016/j.humpath.2006.12.012 (2007).

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

34 Liang, C.-Y., Wang, L.-J., Chen, C.-P., Chen, L.-F., Chen, Y.-H. & Chen, H. GCM1 Regulation of the Expression of Syncytin 2 and Its Cognate Receptor MFSD2A in Human Placenta1. *Biology of Reproduction* **83**, 387-395, doi:10.1095/biolreprod.110.083915 (2010).

35 Renaud, S. J., Chakraborty, D., Mason, C. W., Rumi, M. A. K., Vivian, J. L. & Soares, M. J. OVO-like 1 regulates progenitor cell fate in human trophoblast development. *Proceedings of the National Academy of Sciences* **112**, E6175-E6184, doi:10.1073/pnas.1507397112 (2015).

36 Knerr, I., Schubert, S. W., Wich, C., Amann, K., Aigner, T., Vogler, T., Jung, R., Dötsch, J., Rascher, W. & Hashemolhosseini, S. Stimulation of GCMa and syncytin via cAMP mediated PKA signaling in human trophoblastic cells under normoxic and hypoxic conditions. *FEBS Letters* **579**, 3991-3998, doi:10.1016/j.febslet.2005.06.029 (2005).

37 Yoshie, M., Kaneyama, K., Kusama, K., Higuma, C., Nishi, H., Isaka, K. & Tamura, K. Possible role of the exchange protein directly activated by cyclic AMP (Epac) in the cyclic AMP-dependent functional differentiation and syncytialization of human placental BeWo cells. *Human Reproduction* **25**, 2229-2238, doi:10.1093/humrep/deq190 (2010).

38 De Rooij, J., Zwartkruis, F. J. T., Verheijen, M. H. G., Cool, R. H., Nijman, S. M. B., Wittinghofer, A. & Bos, J. L. Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature* **396**, 474-477, doi:10.1038/24884 (1998).

39 Kawasaki, H., Springett, G. M., Mochizuki, N., Toki, S., Nakaya, M., Matsuda, M., Housman, D. E. & Graybiel, A. M. A Family of cAMP-Binding Proteins That Directly Activate Rap1. *Science* **282**, 2275-2279, doi:10.1126/science.282.5397.2275 (1998).

40 Mi, S., Lee, X., Li, X.-P., Veldman, G. M., Finnerty, H., Racie, L., Lavallie, E., Tang, X.-Y., Edouard, P., Howes, S., Keith, J. C. & McCoy, J. M. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785-789, doi:10.1038/35001608 (2000).

41 Blond, J. L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., Mandrand, B., Mallet, F. & Cosset, F. L. An Envelope Glycoprotein of the Human Endogenous Retrovirus HERV-W Is Expressed in the Human Placenta and Fuses Cells Expressing the Type D Mammalian Retrovirus Receptor. *Journal of Virology* **74**, 3321-3329, doi:10.1128/jvi.74.7.3321-3329.2000 (2000).

42 Esnault, C., Priet, S., Ribet, D., Vernochet, C., Bruls, T., Lavialle, C., Weissenbach, J. & Heidmann, T. A placenta-specific receptor for the fusogenic, endogenous retrovirus-derived, human syncytin-2. *Proceedings of the National Academy of Sciences* **105**, 17532-17537, doi:10.1073/pnas.0807413105 (2008).

43 Tuckey, R. C. Progesterone synthesis by the human placenta. *Placenta* **26**, 273-281, doi:10.1016/j.placenta.2004.06.012 (2005).

44 Pidoux, G., Gerbaud, P., Tsatsaris, V., Marpeau, O., Ferreira, F., Meduri, G., Guibourdenche, J., Badet, J., Evain-Brion, D. & Frendo, J.-L. Biochemical characterization and modulation of LH/CG-receptor during human trophoblast differentiation. *Journal of Cellular Physiology* **212**, 26-35, doi:10.1002/jcp.20995 (2007).

45 Q. Shi, Z. L., C. Rao, J. Lin. Novel role of human chorionic gonadotropin in differentiation of human cytotrophoblasts. *Endocrinology* **132 3**, 1387-1395 (1993).

46 Xiaoyin Lu, R. W., Cheng Zhu, ..., Yan Gu, James C. Cross, Hongmei Wang. Fine-Tuned and Cell-Cycle-Restricted Expression of Fusogenic Protein Syncytin-2 Maintains Functional Placental Syncytia. *Cell Reports*, doi:10.1016/j.celrep.2017.10.019 (2017).

47 Zhang, Y., Le, T., Grabau, R., Mohseni, Z., Kim, H., Natale, D. R., Feng, L., Pan, H. & Yang, H. TMEM16F phospholipid scramblase mediates trophoblast fusion and placental development. *Science Advances* **6**, eaba0310, doi:10.1126/sciadv.aba0310 (2020).

48 Meinhardt, G., Haider, S., Kunihs, V., Saleh, L., Pollheimer, J., Fiala, C., Hetey, S., Feher, Z., Szilagyi, A., Than, N. G. & Knöfler, M. Pivotal role of the transcriptional co-activator YAP in trophoblast stemness of the developing human placenta. *Proceedings of the National Academy of Sciences* **117**, 13562-13570, doi:10.1073/pnas.2002630117 (2020).

49 Brass, A. L., Huang, I. C., Benita, Y., John, S. P., Krishnan, M. N., Feeley, E. M., Ryan, B. J., Weyer, J. L., Van Der Weyden, L., Fikrig, E., Adams, D. J., Xavier, R. J., Farzan, M. & Elledge, S. J. The IFITM Proteins Mediate Cellular Resistance to Influenza A H1N1 Virus, West Nile Virus, and Dengue Virus. *Cell* **139**, 1243-1254, doi:10.1016/j.cell.2009.12.017 (2009).

50 Julian Buchrieser, S. v. A. D., Thérèse Couderc, Quentin Nevers, Olivier Disson, Caroline Manet, Daniel A. Donahue, Françoise Porrot, Kenzo-Hugo Hillion, Emeline Perthame, Marlene V. Arroyo, Sylvie Souquere, Katinka Ruigrok, Anne Dupressoir, Thierry Heidmann, Xavier Montagutelli, Thierry Fournier, Marc Lecuit, Olivier Schwartz. IFITM proteins inhibit placental syncytiotrophoblast formation and promote fetal demise. *Science*, doi:10.1126/science.aaw7733 (2019).

51 Pijnenborg, R., Dixon, G., Robertson, W. B. & Brosens, I. Trophoblastic invasion of human decidua from 8 to 18 weeks of pregnancy. *Placenta* **1**, 3-19, doi:10.1016/s0143-4004(80)80012-9 (1980).

52 Hustin, J. & Schaaps, J.-P. Echocardiograhic and anatomic studies of the maternotrophoblastic border during the first trimester of pregnancy. *American Journal of Obstetrics and Gynecology* **157**, 162-168, doi:10.1016/s0002-9378(87)80371-x (1987).

53 Wakeland, A. K., Soncin, F., Moretto-Zita, M., Chang, C.-W., Horii, M., Pizzo, D., Nelson, K. K., Laurent, L. C. & Parast, M. M. Hypoxia Directs Human Extravillous Trophoblast Differentiation in a Hypoxia-Inducible Factor–Dependent Manner. *The American Journal of Pathology* **187**, 767-780, doi:10.1016/j.ajpath.2016.11.018 (2017).

54 Haider, S., Meinhardt, G., Saleh, L., Fiala, C., Pollheimer, J. & Knöfler, M. Notch1 controls development of the extravillous trophoblast lineage in the human placenta. *Proceedings of the National Academy of Sciences* **113**, E7710-E7719, doi:10.1073/pnas.1612335113 (2016).

55 Marzioni, D. Expression of ZO-1 and occludin in normal human placenta and in hydatidiform moles. *Molecular Human Reproduction* **7**, 279-285, doi:10.1093/molehr/7.3.279 (2001).

56 Fisher, S. J., Cui, T. Y., Zhang, L., Hartman, L., Grahl, K., Zhang, G. Y., Tarpey, J. & Damsky, C. H. Adhesive and degradative properties of human placental cytotrophoblast cells in vitro. *Journal of Cell Biology* **109**, 891-902, doi:10.1083/jcb.109.2.891 (1989).

57 Zhou, Y., Fisher, S. J., Janatpour, M., Genbacev, O., Dejana, E., Wheelock, M. & Damsky, C. H. Human cytotrophoblasts adopt a vascular phenotype as they differentiate. A strategy for successful endovascular invasion? *The Journal of Clinical Investigation* **99**, 2139-2151, doi:10.1172/JCI119387 (1997).

58 Vićovac, L. & Aplin, J. D. Epithelial-Mesenchymal Transition during Trophoblast Differentiation. *Cells Tissues Organs* **156**, 202-216, doi:10.1159/000147847 (1996).

59 E. Davies, J., Pollheimer, J., Yong, H. E. J., Kokkinos, M. I., Kalionis, B., Knöfler, M. & Murthi, P. Epithelial-mesenchymal transition during extravillous trophoblast differentiation. *Cell Adhesion & Migration* **10**, 310-321, doi:10.1080/19336918.2016.1170258 (2016).

60 Ng, Y. H., Zhu, H. & Leung, P. C. K. Twist Modulates Human Trophoblastic Cell Invasion via Regulation of N-Cadherin. *Endocrinology* **153**, 925-936, doi:10.1210/en.2011-1488 (2012).

61 Damsky, C. H., Librach, C., Lim, K. H., Fitzgerald, M. L., McMaster, M. T., Janatpour, M., Zhou, Y., Logan, S. K. & Fisher, S. J. Integrin switching regulates normal trophoblast invasion. *Development* **120**, 3657 (1994).

62 Damsky, C. H., Fitzgerald, M. L. & Fisher, S. J. Distribution patterns of extracellular matrix components and adhesion receptors are intricately modulated during first trimester cytotrophoblast differentiation along the invasive pathway, in vivo. *Journal of Clinical Investigation* **89**, 210-222, doi:10.1172/jci115565 (1992).

63 King, A., Boocock, C., Sharkey, A. M., Gardner, L., Beretta, A., Siccardi, A. G. & Yung Wai, L. Evidence for the expression of HLA-C class I mRNA and protein by human first trimester

trophoblast. *Journal of Reproductive Immunology* **31**, 232-233, doi:https://doi.org/10.1016/0165-0378(96)87783-7 (1996).

64 Kovats, S., Main, E. K., Librach, C., Stubblebine, M., Fisher, S. J. & DeMars, R. A class I antigen, HLA-G, expressed in human trophoblasts. *Science* **248**, 220, doi:10.1126/science.2326636 (1990).

65 Chang, W.-L., Liu, Y.-W., Dang, Y.-L., Jiang, X.-X., Xu, H., Huang, X., Wang, Y.-L., Wang, H., Zhu, C., Xue, L.-Q., Lin, H.-Y., Meng, W. & Wang, H. PLAC8, a new marker for human interstitial extravillous trophoblast cells, promotes their invasion and migration. *Development* **145**, dev148932, doi:10.1242/dev.148932 (2018).

66 Burrows, T. D., King, A. & Loke, Y. W. Expression of adhesion molecules by endovascular trophoblast and decidual endothelial cells: Implications for vascular invasion during implantation. *Placenta* **15**, 21-33, doi:https://doi.org/10.1016/S0143-4004(05)80233-4 (1994).

67 Okae, H., Toh, H., Sato, T., Hiura, H., Takahashi, S., Shirane, K., Kabayama, Y., Suyama, M., Sasaki, H. & Arima, T. Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell* **22**, 50-63 e56, doi:10.1016/j.stem.2017.11.004 (2018).

68 Cheryl, Gardner, L., Turco, M., Zhao, N., Matthew, Coleman, N., Rossant, J., Hemberger, M. & Moffett, A. What Is Trophoblast? A Combination of Criteria Define Human First-Trimester Trophoblast. *Stem Cell Reports* **6**, 257-272, doi:10.1016/j.stemcr.2016.01.006 (2016).

69 Xu, R. H., Chen, X., Li, D. S., Li, R., Addicks, G. C., Glennon, C., Zwaka, T. P. & Thomson, J. A. BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol* **20**, 1261-1264, doi:10.1038/nbt761 (2002).

70 Amita, M., Adachi, K., Alexenko, A. P., Sinha, S., Schust, D. J., Schulz, L. C., Roberts, R. M. & Ezashi, T. Complete and unidirectional conversion of human embryonic stem cells to trophoblast by BMP4. *Proceedings of the National Academy of Sciences* **110**, E1212, doi:10.1073/pnas.1303094110 (2013).

71 Xiaodong Liu, J. F. O., Fernando J. Rossello, Jia Ping Tan, Kathryn C. Davidson, Daniela S. Valdes, Jan Schröder, Yu B. Y. Sun, Joseph Chen, Anja S. Knaupp, Guizhi Sun, Hun S. Chy, Ziyi Huang, Jahnvi Pflueger, Jaber Firas, Vincent Tano, Sam Buckberry, Jacob M. Paynter, Michael R. Larcombe, Daniel Poppe, Xin Yi Choo, Carmel M. O'Brien, William A. Pastor, Di Chen, Anna L. Leichter, Haroon Naeem, Pratibha Tripathi, Partha P. Das, Alexandra Grubman, David R. Powell, Andrew L. Laslett, Laurent David, Susan K. Nilsson, Amander T. Clark, Ryan Lister, Christian M. Nefzger, Luciano G. Martelotto, Owen J. L. Rackham & Jose M. Polo Reprogramming roadmap reveals route to human induced trophoblast stem cell. *Nature*, doi:10.1038/s41586-020-2734-6 (2020).

72 Dong, C., Beltcheva, M., Gontarz, P., Zhang, B., Popli, P., Fischer, L. A., Khan, S. A., Park, K. M., Yoon, E. J., Xing, X., Kommagani, R., Wang, T., Solnica-Krezel, L. & Theunissen, T. W. Derivation of trophoblast stem cells from naive human pluripotent stem cells. *Elife* **9**, doi:10.7554/eLife.52504 (2020).

73 Castel, G., Meistermann, D., Bretin, B., Firmin, J., Blin, J., Loubersac, S., Bruneau, A., Chevolleau, S., Kilens, S., Chariau, C., Gaignerie, A., Francheteau, Q., Kagawa, H., Charpentier, E., Flippe, L., François--Campion, V., Haider, S., Dietrich, B., Knöfler, M., Arima, T., Bourdon, J., Rivron, N., Masson, D., Fournier, T., Okae, H., Fréour, T. & David, L. Induction of Human Trophoblast Stem Cells from Somatic Cells and Pluripotent Stem Cells. *Cell Reports* **33**, 108419, doi:10.1016/j.celrep.2020.108419 (2020).

74 Haider, S., Meinhardt, G., Saleh, L., Kunihs, V., Gamperl, M., Kaindl, U., Ellinger, A., Burkard, T. R., Fiala, C., Pollheimer, J., Mendjan, S., Latos, P. A. & Knöfler, M. Self-Renewing Trophoblast Organoids Recapitulate the Developmental Program of the Early Human Placenta. *Stem Cell Reports* **11**, 537-551, doi:10.1016/j.stemcr.2018.07.004 (2018).

75 Turco, M. Y., Gardner, L., Kay, R. G., Hamilton, R. S., Prater, M., Hollinshead, M. S., McWhinnie, A., Esposito, L., Fernando, R., Skelton, H., Reimann, F., Gribble, F. M., Sharkey, A., Marsh, S. G. E., O'Rahilly, S., Hemberger, M., Burton, G. J. & Moffett, A. Trophoblast organoids as a model for maternal-fetal interactions during human placentation. *Nature* **564**, 263-267, doi:10.1038/s41586-018-0753-3 (2018).

76 Turco, M. Y., Gardner, L., Hughes, J., Cindrova-Davies, T., Gomez, M. J., Farrell, L., Hollinshead, M., Marsh, S. G. E., Brosens, J. J., Critchley, H. O., Simons, B. D., Hemberger, M., Koo, B. K., Moffett, A. & Burton, G. J. Long-term, hormone-responsive organoid cultures of human endometrium in a chemically defined medium. *Nat Cell Biol* **19**, 568-577, doi:10.1038/ncb3516 (2017).

77 Rivron, N., Pera, M., Rossant, J., Martinez Arias, A., Zernicka-Goetz, M., Fu, J., van den Brink, S., Bredenoord, A., Dondorp, W., de Wert, G., Hyun, I., Munsie, M. & Isasi, R. Debate ethics of embryo models from stem cells. *Nature* **564**, 183-185, doi:10.1038/d41586-018-07663-9 (2018).

78 Liu, Y., Fan, X., Wang, R., Lu, X., Dang, Y.-L., Wang, H., Lin, H.-Y., Zhu, C., Ge, H., Cross, J. C. & Wang, H. Single-cell RNA-seq reveals the diversity of trophoblast subtypes and patterns of differentiation in the human placenta. *Cell Research* **28**, 819-832, doi:10.1038/s41422-018-0066-y (2018).

79 Vento-Tormo, R., Efremova, M., Botting, R. A., Turco, M. Y., Vento-Tormo, M., Meyer, K. B., Park, J.-E., Stephenson, E., Polański, K., Goncalves, A., Gardner, L., Holmqvist, S., Henriksson, J., Zou, A., Sharkey, A. M., Millar, B., Innes, B., Wood, L., Wilbrey-Clark, A., Payne, R. P., Ivarsson, M. A., Lisgo, S., Filby, A., Rowitch, D. H., Bulmer, J. N., Wright, G. J., Stubbington, M. J. T., Haniffa, M., Moffett, A. & Teichmann, S. A. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347-353, doi:10.1038/s41586-018-0698-6 (2018).

80 Tsang, J. C. H., Vong, J. S. L., Ji, L., Poon, L. C. Y., Jiang, P., Lui, K. O., Ni, Y. B., To, K. F., Cheng, Y. K. Y., Chiu, R. W. K. & Lo, Y. M. D. Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proc Natl Acad Sci U S A* **114**, E7786-E7795, doi:10.1073/pnas.1710470114 (2017).

81 Moffett-King, A. Natural killer cells and pregnancy. *Nature Reviews Immunology* **2**, 656-663, doi:10.1038/nri886 (2002).

82 Shao, X., Cao, G., Chen, D., Liu, J., Yu, B., Liu, M., Li, Y.-X., Cao, B., Sadovsky, Y. & Wang, Y.-L. Placental trophoblast syncytialization potentiates macropinocytosis via mTOR signaling to adapt to reduced amino acid supply. *Proceedings of the National Academy of Sciences* **118**, e2017092118, doi:10.1073/pnas.2017092118 (2021).

83 Perez-Garcia, V., Fineberg, E., Wilson, R., Murray, A., Mazzeo, C. I., Tudor, C., Sienerth, A., White, J. K., Tuck, E., Ryder, E. J., Gleeson, D., Siragher, E., Wardle-Jones, H., Staudt, N., Wali, N., Collins, J., Geyer, S., Busch-Nentwich, E. M., Galli, A., Smith, J. C., Robertson, E., Adams, D. J., Weninger, W. J., Mohun, T. & Hemberger, M. Placentation defects are highly prevalent in embryonic lethal mouse mutants. *Nature* **555**, 463-468, doi:10.1038/nature26002 (2018).

84 Tranquilli, A. L., Dekker, G., Magee, L., Roberts, J., Sibai, B. M., Steyn, W., Zeeman, G. G. & Brown, M. A. The classification, diagnosis and management of the hypertensive disorders of pregnancy: A revised statement from the ISSHP. *Pregnancy Hypertens* **4**, 97-104, doi:10.1016/j.preghy.2014.02.001 (2014).

85 Thornton, J. G. & Onwude, J. L. Convulsions in pregnancy in related gorillas. *American Journal of Obstetrics & Gynecology* **167**, 240-241, doi:10.1016/S0002-9378(11)91665-2 (1992).

86 Marshall, S. A., Hannan, N. J., Jelinic, M., Nguyen, T. P. H., Girling, J. E. & Parry, L. J. Animal models of preeclampsia: translational failings and why. *Am J Physiol Regul Integr Comp Physiol* **314**, R499-R508, doi:10.1152/ajpregu.00355.2017 (2018).

87 Robertson, W. B., Brosens, I. & Dixon, H. G. The pathological response of the vessels of the placental bed to hypertensive pregnancy. *The Journal of Pathology and Bacteriology* **93**, 581-592, doi:https://doi.org/10.1002/path.1700930219 (1967).

88 Brosens, I., Robertson, W. B. & Dixon, H. G. The physiological response of the vessels of the placental bed to normal pregnancy. *J Pathol Bacteriol* **93**, 569-579, doi:10.1002/path.1700930218 (1967).

89 Zhou, Y., Damsky, C. H. & Fisher, S. J. Preeclampsia is associated with failure of human cytotrophoblasts to mimic a vascular adhesion phenotype. One cause of defective

endovascular invasion in this syndrome? *J Clin Invest* **99**, 2152-2164, doi:10.1172/jci119388 (1997).

90 Burton, G. J., Woods, A. W., Jauniaux, E. & Kingdom, J. C. Rheological and physiological consequences of conversion of the maternal spiral arteries for uteroplacental blood flow during human pregnancy. *Placenta* **30**, 473-482, doi:10.1016/j.placenta.2009.02.009 (2009).

91 Hung, T. H., Skepper, J. N. & Burton, G. J. In vitro ischemia-reperfusion injury in term human placenta as a model for oxidative stress in pathological pregnancies. *Am J Pathol* **159**, 1031-1043, doi:10.1016/s0002-9440(10)61778-6 (2001).

92 Yung, H. W., Atkinson, D., Campion-Smith, T., Olovsson, M., Charnock-Jones, D. S. & Burton, G. J. Differential activation of placental unfolded protein response pathways implies heterogeneity in causation of early- and late-onset pre-eclampsia. *J Pathol* **234**, 262-276, doi:10.1002/path.4394 (2014).

93 Pillay, P., Maharaj, N., Moodley, J. & Mackraj, I. Placental exosomes and pre-eclampsia: Maternal circulating levels in normal pregnancies and, early and late onset pre-eclamptic pregnancies. *Placenta* **46**, 18-25, doi:10.1016/j.placenta.2016.08.078 (2016).

94 Chen, Y., Huang, Y., Jiang, R. & Teng, Y. Syncytiotrophoblast-derived microparticle shedding in early-onset and late-onset severe pre-eclampsia. *Int J Gynaecol Obstet* **119**, 234-238, doi:10.1016/j.ijgo.2012.07.010 (2012).

95 Germain, A. M., Romanik, M. C., Guerra, I., Solari, S., Reyes, M. S., Johnson, R. J., Price, K., Karumanchi, S. A. & Valdés, G. Endothelial dysfunction: a link among preeclampsia, recurrent pregnancy loss, and future cardiovascular events? *Hypertension* **49**, 90-95, doi:10.1161/01.HYP.0000251522.18094.d4 (2007).

96 Gammill, H. S., Lin, C. & Hubel, C. A. Endothelial progenitor cells and preeclampsia. *Front Biosci* **12**, 2383-2394, doi:10.2741/2240 (2007).

97 Levine, R. J., Maynard, S. E., Qian, C., Lim, K. H., England, L. J., Yu, K. F., Schisterman, E. F., Thadhani, R., Sachs, B. P., Epstein, F. H., Sibai, B. M., Sukhatme, V. P. & Karumanchi, S. A. Circulating angiogenic factors and the risk of preeclampsia. *N Engl J Med* **350**, 672-683, doi:10.1056/NEJMoa031884 (2004).

98 Rajakumar, A., Cerdeira, A. S., Rana, S., Zsengeller, Z., Edmunds, L., Jeyabalan, A., Hubel, C. A., Stillman, I. E., Parikh, S. M. & Karumanchi, S. A. Transcriptionally active syncytial aggregates in the maternal circulation may contribute to circulating soluble fms-like tyrosine kinase 1 in preeclampsia. *Hypertension* **59**, 256-264, doi:10.1161/hypertensionaha.111.182170 (2012).

99 Maynard, S. E., Min, J. Y., Merchan, J., Lim, K. H., Li, J., Mondal, S., Libermann, T. A., Morgan, J. P., Sellke, F. W., Stillman, I. E., Epstein, F. H., Sukhatme, V. P. & Karumanchi, S. A. Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia. *J Clin Invest* **111**, 649-658, doi:10.1172/jci17189 (2003).

100 Ahmad, S. & Ahmed, A. Elevated placental soluble vascular endothelial growth factor receptor-1 inhibits angiogenesis in preeclampsia. *Circ Res* **95**, 884-891, doi:10.1161/01.RES.0000147365.86159.f5 (2004).

101 Lu, F., Longo, M., Tamayo, E., Maner, W., Al-Hendy, A., Anderson, G. D., Hankins, G. D. & Saade, G. R. The effect of over-expression of sFlt-1 on blood pressure and the occurrence of other manifestations of preeclampsia in unrestrained conscious pregnant mice. *Am J Obstet Gynecol* **196**, 396.e391-397; discussion 396.e397, doi:10.1016/j.ajog.2006.12.024 (2007).

102 Li, Z., Zhang, Y., Ying Ma, J., Kapoun, A. M., Shao, Q., Kerr, I., Lam, A., O'Young, G., Sannajust, F., Stathis, P., Schreiner, G., Karumanchi, S. A., Protter, A. A. & Pollitt, N. S. Recombinant vascular endothelial growth factor 121 attenuates hypertension and improves kidney damage in a rat model of preeclampsia. *Hypertension* **50**, 686-692, doi:10.1161/hypertensionaha.107.092098 (2007).

103 Szalai, G., Romero, R., Chaiworapongsa, T., Xu, Y., Wang, B., Ahn, H., Xu, Z., Chiang, P. J., Sundell, B., Wang, R., Jiang, Y., Plazyo, O., Olive, M., Tarca, A. L., Dong, Z., Qureshi, F., Papp, Z., Hassan, S. S., Hernandez-Andrade, E. & Than, N. G. Full-Length Human Placental sFlt-1-e15a Isoform Induces Distinct Maternal Phenotypes of Preeclampsia in Mice. *PLOS ONE* **10**, e0119547, doi:10.1371/journal.pone.0119547 (2015).

104 Rana, S., Powe, C. E., Salahuddin, S., Verlohren, S., Perschel, F. H., Levine, R. J., Lim, K. H., Wenger, J. B., Thadhani, R. & Karumanchi, S. A. Angiogenic factors and the risk of adverse outcomes in women with suspected preeclampsia. *Circulation* **125**, 911-919, doi:10.1161/circulationaha.111.054361 (2012).

105 Verlohren, S., Stepan, H. & Dechend, R. Angiogenic growth factors in the diagnosis and prediction of pre-eclampsia. *Clin Sci (Lond)* **122**, 43-52, doi:10.1042/cs20110097 (2012).

106 Venkatesha, S., Toporsian, M., Lam, C., Hanai, J., Mammoto, T., Kim, Y. M., Bdolah, Y., Lim, K. H., Yuan, H. T., Libermann, T. A., Stillman, I. E., Roberts, D., D'Amore, P. A., Epstein, F. H., Sellke, F. W., Romero, R., Sukhatme, V. P., Letarte, M. & Karumanchi, S. A. Soluble endoglin contributes to the pathogenesis of preeclampsia. *Nat Med* **12**, 642-649, doi:10.1038/nm1429 (2006).

107 Levine, R. J., Lam, C., Qian, C., Yu, K. F., Maynard, S. E., Sachs, B. P., Sibai, B. M., Epstein, F. H., Romero, R., Thadhani, R. & Karumanchi, S. A. Soluble endoglin and other circulating antiangiogenic factors in preeclampsia. *N Engl J Med* **355**, 992-1005, doi:10.1056/NEJMoa055352 (2006).

108 Kammerer, R. & Zimmermann, W. Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families. *BMC Biology* **8**, 12, doi:10.1186/1741-7007-8-12 (2010).

109 Hammarström, S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol* **9**, 67-81, doi:10.1006/scbi.1998.0119 (1999).

110 Kuespert, K., Pils, S. & Hauck, C. R. CEACAMs: their role in physiology and pathophysiology. *Curr Opin Cell Biol* **18**, 565-571, doi:10.1016/j.ceb.2006.08.008 (2006).

111 Mißbach, S., Aleksic, D., Blaschke, L., Hassemer, T., Lee, K. J., Mansfeld, M., Hänske, J., Handler, J. & Kammerer, R. Alternative splicing after gene duplication drives CEACAM1-paralog diversification in the horse. *BMC Evolutionary Biology* **18**, doi:10.1186/s12862-018-1145-x (2018).

112 Kim, W. M., Huang, Y. H., Gandhi, A. & Blumberg, R. S. CEACAM1 structure and function in immunity and its therapeutic implications. *Semin Immunol* **42**, 101296, doi:10.1016/j.smim.2019.101296 (2019).

113 Beauchemin, N. & Arabzadeh, A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer and Metastasis Reviews* **32**, 643-671, doi:10.1007/s10555-013-9444-6 (2013).

114 Tatarinov Iu, S. & Masiukevich, V. N. [Immunochemical identification of a new beta-1-globulin in the serum of pregnant women]. *Biull Eksp Biol Med* **69**, 66-68 (1970).

115 Kammerer, R., Herse, F. & Zimmermann, W. in *Evolutionary Biology* Ch. Chapter 3, 37-53 (2016).

116 Aleksic, D., Blaschke, L., Mißbach, S., Hänske, J., Weiß, W., Handler, J., Zimmermann, W., Cabrera-Sharp, V., Read, J. E., De Mestre, A. M., O'Riordan, R., Moore, T. & Kammerer, R. Convergent evolution of pregnancy-specific glycoproteins in human and horse. *Reproduction* **152**, 171-184, doi:10.1530/rep-16-0236 (2016).

117 Chuong, E. B., Tong, W. & Hoekstra, H. E. Maternal–Fetal Conflict: Rapidly Evolving Proteins in the Rodent Placenta. *Molecular Biology and Evolution* **27**, 1221-1225, doi:10.1093/molbev/msq034 (2010).

118 Chang, C. L., Semyonov, J., Cheng, P. J., Huang, S. Y., Park, J. I., Tsai, H.-J., Lin, C.-Y., Grützner, F., Soong, Y. K., Cai, J. J. & Hsu, S. Y. T. Widespread Divergence of the CEACAM/PSG Genes in Vertebrates and Humans Suggests Sensitivity to Selection. *PLoS ONE* **8**, e61701, doi:10.1371/journal.pone.0061701 (2013).

119 Dumont, B. L. & Eichler, E. E. Signals of historical interlocus gene conversion in human segmental duplications. *PLoS One* **8**, e75949, doi:10.1371/journal.pone.0075949 (2013).

120 Zhao, L., Triche, E. W., Walsh, K. M., Bracken, M. B., Saftlas, A. F., Hoh, J. & Dewan, A. T. Genome-wide association study identifies a maternal copy-number deletion in PSG11 enriched among preeclampsia patients. *BMC Pregnancy Childbirth* **12**, 61, doi:10.1186/1471-2393-12-61 (2012).

121 Zebhauser, R., Kammerer, R., Eisenried, A., McLellan, A., Moore, T. & Zimmermann, W. Identification of a novel group of evolutionarily conserved members within the rapidly diverging murine Cea family. *Genomics* **86**, 566-580, doi:10.1016/j.ygeno.2005.07.008 (2005).

122 McLellan, A. S., Zimmermann, W. & Moore, T. Conservation of pregnancy-specific glycoprotein (PSG) N domains following independent expansions of the gene families in rodents and primates. *BMC Evol Biol* **5**, 39, doi:10.1186/1471-2148-5-39 (2005).

123 H Chen, W. Y. C., C L Chen, B C Mansfield and J Y Chou. The Carboxyl-terminal Domain of the Human Pregnancy-specific
Glycoprotein Specifies Intracellular'Retention and Stabili.pdf. *J.Biol.Chem.* (1993).

124 Chan, W.-Y., Zheng, Q.-X., McMahon, J. & Teasel, L. Characterization of new members of the pregnancy-specific ?1-glycoprotein family. *Molecular and Cellular Biochemistry* **106**, doi:10.1007/bf00230182 (1991).

125 Qiao-Xi Zheng, L. A. T., W. Lesley Shupert, and Wai-Yee Chan. Characterization of cDNAs of the human pregnancy-specific-glycoproteinfamily, a new subfamily of the immunoglobulin gene superfamily. *Biochemistry*, doi:10.1021/bi00463a030, (1990).

126 Mendoza, M., Lu, D., Ballesteros, A., Blois, S. M., Abernathy, K., Feng, C., Dimitroff, C. J., Zmuda, J., Panico, M., Dell, A., Vasta, G. R., Haslam, S. M. & Dveksler, G. Glycan characterization of pregnancy-specific glycoprotein 1 and its identification as a novel Galectin-1 ligand. *Glycobiology*, doi:10.1093/glycob/cwaa034 (2020).

127 Lee, J. N., Grudzinskas, J. G. & Chard, T. Circulating levels of pregnancy proteins in early and late pregnancy in relation to placental tissue concentration. *Br J Obstet Gynaecol* **86**, 888-890, doi:10.1111/j.1471-0528.1979.tb10717.x (1979).

128 Lin, T. M., Halbert, S. P. & Spellacy, W. N. Measurement of pregnancy-associated plasma proteins during human gestation. *J Clin Invest* **54**, 576-582, doi:10.1172/jci107794 (1974).

129 Towler, C. M., Horne, C. H., Jandial, V., Campbell, D. M. & MacGillivray, I. Plasma levels of pregnancy-specific beta 1-glycoprotein in complicated pregnancies. *Br J Obstet Gynaecol* **84**, 258-263, doi:10.1111/j.1471-0528.1977.tb12573.x (1977).

130 Frängsmyr, L., Israelsson, A., Teglund, S., Matsunaga, T. & Hammarström, S. Evolution of the carcinoembryonic antigen family. structures of CGM9, CGM11 and pregnancy-specific glycoprotein promoters. *Tumour Biol* **21**, 63-81, doi:10.1159/000030112 (2000).

131 Panzetta-Dutari, G. M., Bocco, J. L., Reimund, B., Flury, A. & Patrito, L. C. Nucleotide sequence of a pregnancy-specific beta 1 glycoprotein gene family member. Identification of a functional promoter region and several putative regulatory sequences. *Mol Biol Rep* **16**, 255-262, doi:10.1007/bf00419665 (1992).

132 Camolotto, S., Racca, A., Rena, V., Nores, R., Patrito, L. C., Genti-Raimondi, S. & Panzetta-Dutari, G. M. Expression and transcriptional regulation of individual pregnancy-specific glycoprotein genes in differentiating trophoblast cells. *Placenta* **31**, 312-319, doi:10.1016/j.placenta.2010.01.004 (2010).

133 Nores, R., Blanchon, L., López-Díaz, F., Bocco, J. L., Patrito, L. C., Sapin, V. & Panzetta-Dutari, G. M. Transcriptional control of the human pregnancy-specific glycoprotein 5 gene is dependent on two GT-boxes recognized by the ubiquitous specificity protein 1 (Sp1) transcription factor. *Placenta* **25**, 9-19, doi:10.1016/s0143-4004(03)00213-3 (2004).

134 Blanchon, L., Nores, R., Gallot, D., Marceau, G., Borel, V., Yang, V. W., Bocco, J. L., Lemery, D., Panzetta-Dutari, G. & Sapin, V. Activation of the human pregnancy-specific glycoprotein PSG-5 promoter by KLF4 and Sp1. *Biochem Biophys Res Commun* **343**, 745-753, doi:10.1016/j.bbrc.2006.03.032 (2006).

135 Lopez-Diaz, F., Nores, R., Panzetta-Dutari, G., Slavin, D., Prieto, C., Koritschoner, N. P. & Bocco, J. L. RXRalpha regulates the pregnancy-specific glycoprotein 5 gene transcription through a functional retinoic acid responsive element. *Placenta* **28**, 898-906, doi:10.1016/j.placenta.2007.03.002 (2007).

136 Yan, J., Tanaka, S., Oda, M., Makino, T., Ohgane, J. & Shiota, K. Retinoic acid promotes differentiation of trophoblast stem cells to a giant cell fate. *Dev Biol* **235**, 422-432, doi:10.1006/dbio.2001.0300 (2001).

137 Sapin, V., Dollé, P., Hindelang, C., Kastner, P. & Chambon, P. Defects of the chorioallantoic placenta in mouse RXRalpha null fetuses. *Dev Biol* **191**, 29-41, doi:10.1006/dbio.1997.8687 (1997).

138 Chamberlin, M. E., Lei, K. J. & Chou, J. Y. Subtle differences in human pregnancy-specific glycoprotein gene promoters allow for differential expression. *J Biol Chem* **269**, 17152-17159 (1994).

139 Camolotto, S. A., Racca, A. C., Ridano, M. E., Genti-Raimondi, S. & Panzetta-Dutari, G. M. PSG gene expression is up-regulated by lysine acetylation involving histone and nonhistone proteins. *PloS one* **8**, e55992-e55992, doi:10.1371/journal.pone.0055992 (2013).

140 Jeong, D. S., Kim, M. H. & Lee, J. Y. Depletion of CTCF disrupts PSG gene expression in the human trophoblast cell line Swan 71. *FEBS Open Bio*, doi:10.1002/2211-5463.13087 (2021).

141 Liu, Y. Y., Zhang, S., Yu, T. J., Zhang, F. L., Yang, F., Huang, Y. N., Ma, D., Liu, G. Y., Shao, Z. M. & Li, D. Q. Pregnancy-specific glycoprotein 9 acts as both a transcriptional target and a regulator of the canonical TGF-β/Smad signaling to drive breast cancer progression. *Clinical and Translational Medicine* **10**, doi:10.1002/ctm2.245 (2020).

142 Blois, S. M., Sulkowski, G., Tirado-González, I., Warren, J., Freitag, N., Klapp, B. F., Rifkin, D., Fuss, I., Strober, W. & Dveksler, G. S. Pregnancy-specific glycoprotein 1 (PSG1) activates TGF-β and prevents dextran sodium sulfate (DSS)-induced colitis in mice. *Mucosal Immunology* **7**, 348-358, doi:10.1038/mi.2013.53 (2014).

143 Graham, C. H., Lysiak, J. J., McCrae, K. R. & Lala, P. K. Localization of transforming growth factor-beta at the human fetal-maternal interface: role in trophoblast growth and differentiation. *Biol Reprod* **46**, 561-572, doi:10.1095/biolreprod46.4.561 (1992).

144 Jones, R. L., Stoikos, C., Findlay, J. K. & Salamonsen, L. A. TGF-beta superfamily expression and actions in the endometrium and placenta. *Reproduction* **132**, 217-232, doi:10.1530/rep.1.01076 (2006).

145 Zhao, M. R., Qiu, W., Li, Y. X., Zhang, Z. B., Li, D. & Wang, Y. L. Dual effect of transforming growth factor beta1 on cell adhesion and invasion in human placenta trophoblast cells. *Reproduction* **132**, 333-341, doi:10.1530/rep.1.01112 (2006).

146 Stoikos, C. J., Harrison, C. A., Salamonsen, L. A. & Dimitriadis, E. A distinct cohort of the TGFbeta superfamily members expressed in human endometrium regulate decidualization. *Hum Reprod* **23**, 1447-1456, doi:10.1093/humrep/den110 (2008).

147 Lisboa, F. A., Warren, J., Sulkowski, G., Aparicio, M., David, G., Zudaire, E. & Dveksler, G. S. Pregnancy-specific Glycoprotein 1 Induces Endothelial Tubulogenesis through Interaction with Cell Surface Proteoglycans. *Journal of Biological Chemistry* **286**, 7577-7586, doi:10.1074/jbc.m110.161810 (2011).

148 Shanley, D. K., Kiely, P. A., Golla, K., Allen, S., Martin, K., O'Riordan, R. T., Ball, M., Aplin, J. D., Singer, B. B., Caplice, N., Moran, N. & Moore, T. Pregnancy-Specific Glycoproteins Bind Integrin αIIbβ3 and Inhibit the Platelet—Fibrinogen Interaction. *PLoS ONE* **8**, e57491, doi:10.1371/journal.pone.0057491 (2013).

149 Blois, S. M., Dveksler, G., Vasta, G. R., Freitag, N., Blanchard, V. & Barrientos, G. Pregnancy Galectinology: Insights Into a Complex Network of Glycan Binding Proteins. *Frontiers in immunology* **10**, 1166-1166, doi:10.3389/fimmu.2019.01166 (2019).

150 Than, N. G., Romero, R., Balogh, A., Karpati, E., Mastrolia, S. A., Staretz-Chacham, O., Hahn, S., Erez, O., Papp, Z. & Kim, C. J. Galectins: Double-edged Swords in the Cross-roads of Pregnancy Complications and Female Reproductive Tract Inflammation and Neoplasia. *J Pathol Transl Med* **49**, 181-208, doi:10.4132/jptm.2015.02.25 (2015).

151 Barrientos, G., Freitag, N., Tirado-González, I., Unverdorben, L., Jeschke, U., Thijssen, V. L. & Blois, S. M. Involvement of galectin-1 in reproduction: past, present and future. *Hum Reprod Update* **20**, 175-193, doi:10.1093/humupd/dmt040 (2014).

152 Zambuto, S. G., Rattila, S., Dveksler, G. & Harley, B. A. C. The role of pregnancy-specific glycoproteins on trophoblast motility in three-dimensional gelatin hydrogels. *bioRxiv*, 2020.2009.2025.314195, doi:10.1101/2020.09.25.314195 (2020).

153 Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18**, 71-86, doi:10.1038/nrg.2016.139 (2017).

154 Wells, J. N. & Feschotte, C. A Field Guide to Eukaryotic Transposable Elements. *Annual Review of Genetics* **54**, doi:10.1146/annurev-genet-040620-022145 (2020).

155 Burns, K. H. Our Conflict with Transposable Elements and Its Implications for Human Disease. *Annual Review of Pathology: Mechanisms of Disease* **15**, 51-70, doi:10.1146/annurev-pathmechdis-012419-032633 (2020).

156 McClintock, B. Intranuclear systems controlling gene action and mutation. *Brookhaven Symp Biol*, 58-74 (1956).

157 Lanciano, S. & Cristofari, G. Measuring and interpreting transposable element expression. *Nature Reviews Genetics*, doi:10.1038/s41576-020-0251-y (2020).

158 Payer, L. M. & Burns, K. H. Transposable elements in human genetic disease. *Nature Reviews Genetics*, doi:10.1038/s41576-019-0165-8 (2019).

159 Finnegan, D. J. Eukaryotic transposable elements and genome evolution. *Trends Genet* **5**, 103-107, doi:10.1016/0168-9525(89)90039-5 (1989).

160 Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**, 691-703, doi:10.1038/nrg2640 (2009).

161 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C. & Baldwin, J. e. a. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, doi:10.1038/35057062 (2001).

162 Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V. & Kazazian, H. H., Jr. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* **100**, 5280-5285, doi:10.1073/pnas.0831042100 (2003).

163 Batzer, M. A. & Deininger, P. L. Alu repeats and human genomic diversity. *Nat Rev Genet* **3**, 370-379, doi:10.1038/nrg798 (2002).

164 Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35**, 41-48, doi:10.1038/ng1223 (2003).

165 Wilhelm, M. & Wilhelm, F. X. Reverse transcription of retroviruses and LTR retrotransposons. *Cell Mol Life Sci* **58**, 1246-1262, doi:10.1007/pl00000937 (2001).

166 Mager, D. L. & Stoye, J. P. Mammalian Endogenous Retroviruses. *Microbiol Spectr* **3**, Mdna3-0009-2014, doi:10.1128/microbiolspec.MDNA3-0009-2014 (2015).

167 Eickbush, T. H. & Malik, H. S. in *Mobile DNA II* (American Society of Microbiology, 2002).

168 Johnson, W. E. Origins and evolutionary consequences of ancient endogenous retroviruses. *Nature Reviews Microbiology* **17**, 355-370, doi:10.1038/s41579-019-0189-2 (2019).

169 Magiorkinis, G., Gifford, R. J., Katzourakis, A., De Ranter, J. & Belshaw, R. Env-less endogenous retroviruses are genomic superspreaders. *Proc Natl Acad Sci U S A* **109**, 7385-7390, doi:10.1073/pnas.1200913109 (2012).

170 Frank, J. A. & Feschotte, C. Co-option of endogenous viral sequences for host cell function. *Current Opinion in Virology* **25**, 81-89, doi:https://doi.org/10.1016/j.coviro.2017.07.021 (2017).

171 Johnson, W. E. Endogenous Retroviruses in the Genomics Era. *Annu Rev Virol* **2**, 135-159, doi:10.1146/annurev-virology-100114-054945 (2015).

172 Dupressoir, A., Lavialle, C. & Heidmann, T. From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. *Placenta* **33**, 663-671, doi:10.1016/j.placenta.2012.05.005 (2012).

173 Chuong, E. B. Retroviruses facilitate the rapid evolution of the mammalian placenta. *BioEssays*, n/a-n/a, doi:10.1002/bies.201300059 (2013).

174 Haig, D. Retroviruses and the Placenta. *Current Biology* **22**, R609-R613, doi:10.1016/j.cub.2012.06.002 (2012).

175 Rawn, S. M. & Cross, J. C. The Evolution, Regulation, and Function of Placenta-Specific Genes. *Annual Review of Cell and Developmental Biology* **24**, 159-181, doi:10.1146/annurev.cellbio.24.110707.175418 (2008).

176 Aiewsakun, P. & Katzourakis, A. Marine origin of retroviruses in the early Palaeozoic Era. *Nature Communications* **8**, 13954, doi:10.1038/ncomms13954 (2017).

177 Geis, F. K. & Goff, S. P. Silencing and Transcriptional Regulation of Endogenous Retroviruses: An Overview. *Viruses* **12**, 884, doi:10.3390/v12080884 (2020).

178 Copeland, N. G., Hutchison, K. W. & Jenkins, N. A. Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral LTRs. *Cell* **33**, 379-387, doi:10.1016/0092-8674(83)90419-1 (1983).
179 Chapman, V., Forrester, L., Sanford, J., Hastie, N. & Rossant, J. Cell lineage-specific undermethylation of mouse repetitive DNA. *Nature* **307**, 284-286, doi:10.1038/307284a0 (1984).
180 Golding, M. C., Zhang, L. & Mann, M. R. W. Multiple Epigenetic Modifiers Induce Aggressive Viral Extinction in Extraembryonic Endoderm Stem Cells. *Cell Stem Cell* **6**, 457-467, doi:10.1016/j.stem.2010.03.014 (2010).
181 Zhang, Y., Xiang, Y., Yin, Q., Du, Z., Peng, X., Wang, Q., Fidalgo, M., Xia, W., Li, Y., Zhao, Z.-A., Zhang, W., Ma, J., Xu, F., Wang, J., Li, L. & Xie, W. Dynamic epigenomic landscapes during early lineage specification in mouse embryos. *Nature Genetics* **50**, 96-105, doi:10.1038/s41588-017-0003-x (2018).
182 Cohen, C. J., Rebollo, R., Babovic, S., Dai, E. L., Robinson, W. P. & Mager, D. L. Placenta-specific expression of the interleukin-2 (IL-2) receptor beta subunit from an endogenous retroviral promoter. *J Biol Chem* **286**, 35543-35552, doi:10.1074/jbc.M111.227637 (2011).
183 Reiss, D., Zhang, Y. & Mager, D. L. Widely variable endogenous retroviral methylation levels in human placenta. *Nucleic Acids Res* **35**, 4743-4754, doi:10.1093/nar/gkm455 (2007).
184 Chuong, E. B., Rumi, M. A., Soares, M. J. & Baker, J. C. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* **45**, 325-329, doi:10.1038/ng.2553 (2013).
185 Cosset, F.-L. & Lavillette, D. Cell entry of enveloped viruses. *Adv Genet* **73**, 121-183, doi:10.1016/B978-0-12-380860-8.00004-5 (2011).
186 Blaise, S., de Parseval, N., Bénit, L. & Heidmann, T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc Natl Acad Sci U S A* **100**, 13013-13018, doi:10.1073/pnas.2132646100 (2003).
187 Hunter, E. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 1997).
188 Toudic, C., Elisseeff, X., Lokossou, A. G. & Barbeau, B. in *Human Retrotransposons in Health and Disease* Ch. Chapter 9, 215-238 (2017).
189 Muir, A., Lever, A. M. L. & Moffett, A. Human endogenous retrovirus-W envelope (syncytin) is expressed in both villous and extravillous trophoblast populations. *J Gen Virol* **87**, 2067-2071, doi:10.1099/vir.0.81412-0 (2006).
190 Lavillette, D., Marin, M., Ruggieri, A., Mallet, F. O., Cosset, F. O.-L. C. & Kabat, D. The Envelope Glycoprotein of Human Endogenous Retrovirus Type W Uses a Divergent Family of Amino Acid Transporters/Cell Surface Receptors. *Journal of Virology* **76**, 6442-6452, doi:10.1128/jvi.76.13.6442-6452.2002 (2002).
191 Malassine, A., Blaise, S., Handschuh, K., Lalucque, H., Dupressoir, A., Evain-Brion, D. & Heidmann, T. Expression of the fusogenic HERV-FRD Env glycoprotein (syncytin 2) in human placenta is restricted to villous cytotrophoblastic cells. *Placenta* **28**, 185-191, doi:10.1016/j.placenta.2006.03.001 (2007).
192 Frendo, J.-L., Olivier, D., Cheynet, V., Blond, J.-L., Bouton, O., Vidaud, M., Rabreau, M., Evain-Brion, D. & Mallet, F. Direct Involvement of HERV-W Env Glycoprotein in Human Trophoblast Cell Fusion and Differentiation. *Molecular and Cellular Biology* **23**, 3566, doi:10.1128/MCB.23.10.3566-3574.2003 (2003).
193 Vargas, A., Moreau, J., Landry, S., LeBellego, F., Toufaily, C., Rassart, E., Lafond, J. & Barbeau, B. Syncytin-2 plays an important role in the fusion of human trophoblast cells. *J Mol Biol* **392**, 301-318, doi:10.1016/j.jmb.2009.07.025 (2009).
194 Dupressoir, A., Vernochet, C., Harper, F., Guegan, J., Dessen, P., Pierron, G. & Heidmann, T. A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. *Proceedings of the National Academy of Sciences* **108**, E1164-E1173, doi:10.1073/pnas.1112304108 (2011).
195 Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P. & Heidmann, T. Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic,

endogenous retrovirus-derived, envelope gene. *Proceedings of the National Academy of Sciences* **106**, 12127-12132, doi:10.1073/pnas.0902925106 (2009).

196 Lokossou, A. G., Toudic, C., Nguyen, P. T., Elisseeff, X., Vargas, A., Rassart, E., Lafond, J., Leduc, L., Bourgault, S., Gilbert, C., Scorza, T., Tolosa, J. & Barbeau, B. Endogenous retrovirus-encoded Syncytin-2 contributes to exosome-mediated immunosuppression of T cellsdagger. *Biol Reprod* **102**, 185-198, doi:10.1093/biolre/ioz124 (2020).

197 Tolosa, J. M., Schjenken, J. E., Clifton, V. L., Vargas, A., Barbeau, B., Lowry, P., Maiti, K. & Smith, R. The endogenous retroviral envelope protein syncytin-1 inhibits LPS/PHA-stimulated cytokine responses in human blood and is sorted into placental exosomes. *Placenta* **33**, 933-941, doi:10.1016/j.placenta.2012.08.004 (2012).

198 Sugimoto, J., Sugimoto, M., Bernstein, H., Jinno, Y. & Schust, D. A novel human endogenous retroviral protein inhibits cell-cell fusion. *Scientific Reports* **3**, doi:10.1038/srep01462 (2013).

199 Vargas, A., Toufaily, C., Lebellego, F., Rassart, É., Lafond, J. & Barbeau, B. Reduced Expression of Both Syncytin 1 and Syncytin 2 Correlates With Severity of Preeclampsia. *Reproductive Sciences* **18**, 1085-1091, doi:10.1177/1933719111404608 (2011).

200 Tang, Y., Woodward, B. O., Pastor, L., George, A. M., Petrechko, O., Nouvet, F. J., Haas, D. W., Jiang, G. & Hildreth, J. E. K. Endogenous Retroviral Envelope Syncytin Induces HIV-1 Spreading and Establishes HIV Reservoirs in Placenta. *Cell Reports* **30**, 4528-4539.e4524, doi:10.1016/j.celrep.2020.03.016 (2020).

201 Roberts, R. M., Ezashi, T., Schulz, L. C., Sugimoto, J., Schust, D. J., Khan, T. & Zhou, J. Syncytins expressed in human placental trophoblast. *Placenta*, doi:https://doi.org/10.1016/j.placenta.2021.01.006 (2021).

202 Cornelis, G., Vernochet, C., Carradec, Q., Souquere, S., Mulot, B., Catzeflis, F., Nilsson, M. A., Menzies, B. R., Renfree, M. B., Pierron, G., Zeller, U., Heidmann, O., Dupressoir, A. & Heidmann, T. Retroviral envelope gene captures andsyncytinexaptation for placentation in marsupials. *Proceedings of the National Academy of Sciences* **112**, E487-E496, doi:10.1073/pnas.1417000112 (2015).

203 Carter, A. M. & Enders, A. C. Comparative aspects of trophoblast development and placentation. *Reproductive Biology and Endocrinology* **2**, 46, doi:10.1186/1477-7827-2-46 (2004).

204 Griffith, O. W. & Wagner, G. P. The placenta as a model for understanding the origin and evolution of vertebrate organs. *Nature Ecology & Evolution* **1**, 0072, doi:10.1038/s41559-017-0072 (2017).

205 Cornelis, G., Funk, M., Vernochet, C., Leal, F., Tarazona, O. A., Meurice, G., Heidmann, O., Dupressoir, A., Miralles, A., Ramirez-Pinilla, M. P. & Heidmann, T. An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental Mabuya lizard. *Proceedings of the National Academy of Sciences* **114**, E10991-E11000, doi:10.1073/pnas.1714590114 (2017).

206 Branco, M. R. & Chuong, E. B. Crossroads between transposons and gene regulation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**, 20190330, doi:10.1098/rstb.2019.0330 (2020).

207 Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* **12**, 7-18, doi:10.1038/nrg2905 (2011).

208 Ng, R. K., Dean, W., Dawson, C., Lucifero, D., Madeja, Z., Reik, W. & Hemberger, M. Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat Cell Biol* **10**, 1280-1290, doi:10.1038/ncb1786 (2008).

209 Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R. & Rossant, J. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917-929, doi:10.1016/j.cell.2005.08.040 (2005).

210 Russ, A. P., Wattler, S., Colledge, W. H., Aparicio, S. A., Carlton, M. B., Pearce, J. J., Barton, S. C., Surani, M. A., Ryan, K., Nehls, M. C., Wilson, V. & Evans, M. J. Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404**, 95-99, doi:10.1038/35003601 (2000).

211 Todd, C. D., Deniz, Ö., Taylor, D. & Branco, M. R. Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *eLife* **8**, doi:10.7554/elife.44344 (2019).

212 Ferreira, L. M. R., Meissner, T. B., Mikkelsen, T. S., Mallard, W., O'Donnell, C. W., Tilburgs, T., Gomes, H. A. B., Camahort, R., Sherwood, R. I., Gifford, D. K., Rinn, J. L., Cowan, C. A. & Strominger, J. L. A distant trophoblast-specific enhancer controls HLA-G expression at the maternal–fetal interface. *Proceedings of the National Academy of Sciences* **113**, 5364-5369, doi:10.1073/pnas.1602886113 (2016).

213 Dunn-Fletcher, C. E., Muglia, L. M., Pavlicev, M., Wolf, G., Sun, M. A., Hu, Y. C., Huffman, E., Tumukuntala, S., Thiele, K., Mukherjee, A., Zoubovsky, S., Zhang, X., Swaggart, K. A., Lamm, K. Y. B., Jones, H., Macfarlan, T. S. & Muglia, L. J. Anthropoid primate-specific retroviral element THE1B controls expression of CRH in placenta and alters gestation length. *PLoS Biol* **16**, e2006337, doi:10.1371/journal.pbio.2006337 (2018).

214 Ball, M., McLellan, A., Collins, B., Coadwell, J., Stewart, F. & Moore, T. An abundant placental transcript containing an IAP-LTR is allelic to mouse pregnancy-specific glycoprotein 23 (Psg23): cloning and genetic analysis. *Gene* **325**, 103-113, doi:10.1016/j.gene.2003.10.001 (2004).

215 Lee, J. Y., Ji, Z. & Tian, B. Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes. *Nucleic Acids Res* **36**, 5581-5590, doi:10.1093/nar/gkn540 (2008).

216 Pattillo, R. A. & Gey, G. O. The establishment of a cell line of human hormone-synthesizing trophoblastic cells in vitro. *Cancer Res* **28**, 1231-1236 (1968).

217 Choy, M. Y. & Manyonda, I. T. The phagocytic activity of human first trimester extravillous trophoblast. *Hum Reprod* **13**, 2941-2949, doi:10.1093/humrep/13.10.2941 (1998).

218 Mátés, L., Chuah, M. K. L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D. P., Schmitt, A., Becker, K., Matrai, J., Ma, L., Samara-Kuko, E., Gysemans, C., Pryputniewicz, D., Miskey, C., Fletcher, B., Vandendriessche, T., Ivics, Z. & Izsvák, Z. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nature Genetics* **41**, 753-761, doi:10.1038/ng.343 (2009).

219 Zadora, J., Singh, M., Herse, F., Przybyl, L., Haase, N., Golic, M., Yung, H. W., Huppertz, B., Cartwright, J. E., Whitley, G., Johnsen, G. M., Levi, G., Isbruch, A., Schulz, H., Luft, F. C., Müller, D. N., Staff, A. C., Hurst, L. D., Dechend, R. & Izsvák, Z. Disturbed Placental Imprinting in Preeclampsia Leads to Altered Expression of DLX5, a Human-Specific Early Trophoblast Marker. *Circulation* **136**, 1824-1839, doi:10.1161/circulationaha.117.028110 (2017).

220 Deng, Z., Zhang, S., Gu, S., Ni, X., Zeng, W. & Li, X. Useful Bicistronic Reporter System for Studying Poly(A) Site-Defining cis Elements and Regulation of Alternative Polyadenylation. *Int J Mol Sci* **19**, doi:10.3390/ijms19010279 (2018).

221 Hu, J., Lutz, C. S., Wilusz, J. & Tian, B. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA* **11**, 1485-1493, doi:10.1261/rna.2107305 (2005).

222 Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research* **46**, W242-W245, doi:10.1093/nar/gky354 (2018).

223 Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. & Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).

224 Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).

225 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

226 Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184-2185, doi:10.1093/bioinformatics/bts356 (2012).

227 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

228 Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research* **47**, e47-e47, doi:10.1093/nar/gkz114 (2019).

229 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

230 Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48, doi:10.1186/1471-2105-10-48 (2009).

231 Angelova, M., Machado, H. L., Swan, K. F., Morris, C. & Sullivan, D. E. Extravillous Trophoblast Migration and Invasion Assay. *Bio Protoc* **3**, doi:10.21769/bioprotoc.840 (2013).

232 Kwak, Y. T., Muralimanoharan, S., Gogate, A. A. & Mendelson, C. R. Human Trophoblast Differentiation Is Associated With Profound Gene Regulatory and Epigenetic Changes. *Endocrinology* **160**, 2189-2203, doi:10.1210/en.2019-00144 (2019).

233 Domcke, S., Hill, A. J., Daza, R. M., Cao, J., O'Day, D. R., Pliner, H. A., Aldinger, K. A., Pokholok, D., Zhang, F., Milbank, J. H., Zager, M. A., Glass, I. A., Steemers, F. J., Doherty, D., Trapnell, C., Cusanovich, D. A. & Shendure, J. A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612, doi:10.1126/science.aba7612 (2020).

234 Yang, D., Jang, I., Choi, J., Kim, M. S., Lee, A. J., Kim, H., Eom, J., Kim, D., Jung, I. & Lee, B. 3DIV: A 3D-genome Interaction Viewer and database. *Nucleic Acids Res* **46**, D52-D57, doi:10.1093/nar/gkx1017 (2018).

235 Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., Diao, Y., Liang, J., Zhao, H., Lobanenkov, V. V., Ecker, J. R., Thomson, J. A. & Ren, B. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336, doi:10.1038/nature14222 (2015).

236 Jung, I., Schmitt, A., Diao, Y., Lee, A. J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S., Chiang, Z., Kim, C., Masliah, E., Barr, C. L., Li, B., Kuan, S., Kim, D. & Ren, B. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nature Genetics* **51**, 1442-1449, doi:10.1038/s41588-019-0494-8 (2019).

237 Gruber, A. J. & Zavolan, M. Alternative cleavage and polyadenylation in health and disease. *Nature Reviews Genetics* **20**, 599-614, doi:10.1038/s41576-019-0145-z (2019).

238 Tian, B. & Manley, J. L. Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology* **18**, 18-30, doi:10.1038/nrm.2016.116 (2017).

239 Hu, J. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. **11**, 1485-1493, doi:10.1261/rna.2107305 (2005).

240 Black, S. D. & Mould, D. R. Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. *Anal Biochem* **193**, 72-82, doi:10.1016/0003-2697(91)90045-u (1991).

241 Wice, B., Menton, D., Geuze, H. & Schwartz, A. L. Modulators of cyclic AMP metabolism induce syncytiotrophoblast formation in vitro. *Exp Cell Res* **186**, 306-316, doi:10.1016/0014-4827(90)90310-7 (1990).

242 Ringler, G. E. & Strauss, J. F., 3rd. In vitro systems for the study of human placental endocrine function. *Endocr Rev* **11**, 105-123, doi:10.1210/edrv-11-1-105 (1990).

243 Cheng, L. C., Zheng, D., Baljinnyam, E., Sun, F., Ogami, K., Yeung, P. L., Hoque, M., Lu, C.-W., Manley, J. L. & Tian, B. Widespread transcript shortening through alternative polyadenylation in secretory cell differentiation. *Nature Communications* **11**, doi:10.1038/s41467-020-16959-2 (2020).

244 Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology* **38**, 824-844, doi:10.1038/s41587-020-0561-9 (2020).

245 Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
246 Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11, doi:10.1186/s13100-015-0041-9 (2015).
247 Kammerer, R., Popp, T., Singer, B. B., Schlender, J. & Zimmermann, W. Identification of allelic variants of the bovine immune regulatory molecule CEACAM1 implies a pathogen-driven evolution. *Gene* **339**, 99-109, doi:10.1016/j.gene.2004.06.023 (2004).
248 Kammerer, R., Hahn, S., Singer, B. B., Luo, J. S. & von Kleist, S. Biliary glycoprotein (CD66a), a cell adhesion molecule of the immunoglobulin superfamily, on human lymphocytes: structure, expression and involvement in T cell activation. *Eur J Immunol* **28**, 3664-3674, doi:10.1002/(sici)1521-4141(199811)28:11<3664::Aid-immu3664>3.0.Co;2-d (1998).
249 MacDonald, C. C., Wilusz, J. & Shenk, T. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Molecular and cellular biology* **14**, 6647-6654, doi:10.1128/mcb.14.10.6647 (1994).
250 Takagaki, Y., Seipelt, R. L., Peterson, M. L. & Manley, J. L. The Polyadenylation Factor CstF-64 Regulates Alternative Processing of IgM Heavy Chain Pre-mRNA during B Cell Differentiation. *Cell* **87**, 941-952, doi:10.1016/s0092-8674(00)82000-0 (1996).
251 Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J. & Zhang, F. RNA editing with CRISPR-Cas13. *Science* **358**, 1019-1027, doi:10.1126/science.aaq0180 (2017).
252 Toudic, C., Vargas, A., Xiao, Y., St-Pierre, G., Bannert, N., Lafond, J., Rassart, É., Sato, S. & Barbeau, B. Galectin-1 interacts with the human endogenous retroviral envelope protein syncytin-2 and potentiates trophoblast fusion in humans. *The FASEB Journal* **33**, 12873-12887, doi:10.1096/fj.201900107r (2019).
253 West, R. C., Ming, H., Logsdon, D. M., Sun, J., Rajput, S. K., Kile, R. A., Schoolcraft, W. B., Roberts, R. M., Krisher, R. L., Jiang, Z. & Yuan, Y. Dynamics of trophoblast differentiation in peri-implantation–stage human embryos. *Proceedings of the National Academy of Sciences*, 201911362, doi:10.1073/pnas.1911362116 (2019).
254 Park, J., Cha, D. H., Lee, S. J., Kim, Y. N., Kim, Y. H. & Kim, K. P. Discovery of the serum biomarker proteins in severe preeclampsia by proteomic analysis. *Exp Mol Med* **43**, 427-435, doi:10.3858/emm.2011.43.7.047 (2011).
255 DiFederico, E., Genbacev, O. & Fisher, S. J. Preeclampsia is associated with widespread apoptosis of placental cytotrophoblasts within the uterine wall. *Am J Pathol* **155**, 293-301, doi:10.1016/s0002-9440(10)65123-1 (1999).
256 Leung, D. N., Smith, S. C., To, K. F., Sahota, D. S. & Baker, P. N. Increased placental apoptosis in pregnancies complicated by preeclampsia. *Am J Obstet Gynecol* **184**, 1249-1250, doi:10.1067/mob.2001.112906 (2001).
257 Ishihara, N., Matsuo, H., Murakoshi, H., Laoag-Fernandez, J. B., Samoto, T. & Maruo, T. Increased apoptosis in the syncytiotrophoblast in human term placentas complicated by either preeclampsia or intrauterine growth retardation. *Am J Obstet Gynecol* **186**, 158-166, doi:10.1067/mob.2002.119176 (2002).
258 Longtine, M. S., Chen, B., Odibo, A. O., Zhong, Y. & Nelson, D. M. Villous trophoblast apoptosis is elevated and restricted to cytotrophoblasts in pregnancies complicated by preeclampsia, IUGR, or preeclampsia with IUGR. *Placenta* **33**, 352-359, doi:10.1016/j.placenta.2012.01.017 (2012).
259 Munchel, S., Rohrback, S., Randise-Hinchliff, C., Kinnings, S., Deshmukh, S., Alla, N., Tan, C., Kia, A., Greene, G., Leety, L., Rhoa, M., Yeats, S., Saul, M., Chou, J., Bianco, K., O'Shea, K., Bujold, E., Norwitz, E., Wapner, R., Saade, G. & Kaper, F. Circulating transcripts in maternal blood reflect a molecular signature of early-onset preeclampsia. *Science Translational Medicine* **12**, eaaz0131, doi:10.1126/scitranslmed.aaz0131 (2020).
260 Zhao, L., Triche, E. W., Walsh, K. M., Bracken, M. B., Saftlas, A. F., Hoh, J. & Dewan, A. T. Genome-wide association study identifies a maternal copy-number deletion in PSG11

enriched among preeclampsia patients. *BMC Pregnancy and Childbirth* **12**, 61, doi:10.1186/1471-2393-12-61 (2012).

261 Yeung, K. R., Chiu, C. L., Pidsley, R., Makris, A., Hennessy, A. & Lind, J. M. DNA methylation profiles in preeclampsia and healthy control placentas. *American Journal of Physiology-Heart and Circulatory Physiology* **310**, H1295-H1303, doi:10.1152/ajpheart.00958.2015 (2016).

262 Yuen, R. K. C., Peñaherrera, M. S., von Dadelszen, P., McFadden, D. E. & Robinson, W. P. DNA methylation profiling of human placentas reveals promoter hypomethylation of multiple genes in early-onset preeclampsia. *European Journal of Human Genetics* **18**, 1006-1012, doi:10.1038/ejhg.2010.63 (2010).

263 Cruz, J. d. O., Conceição, I. M. C. A., Tosatti, J. A. G., Gomes, K. B. & Luizon, M. R. Global DNA methylation in placental tissues from pregnant with preeclampsia: A systematic review and pathway analysis. *Placenta* **101**, 97-107, doi:https://doi.org/10.1016/j.placenta.2020.09.004 (2020).

264 Meister, S., Hahn, L., Beyer, S., Kuhn, C., Jegen, M., von Schönfeldt, V., Corradini, S., Schulz, C., Kolben, T. M., Hester, A., Appelt, T., Mahner, S., Jeschke, U. & Kolben, T. Epigenetic modification via H3K4me3 and H3K9ac in human placenta is reduced in preeclampsia. *Journal of Reproductive Immunology*, 103287, doi:https://doi.org/10.1016/j.jri.2021.103287 (2021).

265 Sheridan, M. A., Yang, Y., Jain, A., Lyons, A. S., Yang, P., Brahmasani, S. R., Dai, A., Tian, Y., Ellersieck, M. R., Tuteja, G., Schust, D. J., Schulz, L. C., Ezashi, T. & Roberts, R. M. Early onset preeclampsia in a model for human placental trophoblast. *Proceedings of the National Academy of Sciences* **116**, 4336-4345, doi:10.1073/pnas.1816150116 (2019).

7. Statutory declaration

“I, Yuliang Qu, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic: “Humane endogene Retroviren-gesteuerte Expression von schwangerschaftsspezifischem Glykoprotein 9 bei Plazentation und Prä-eklampsie” or “Human Endogenous Retroviruses-Directed Expression of Pregnancy-Specific Glycoprotein 9 in Placentation and Pre-eclampsia”, independently and without the support of third parties, and that I used no other sources and aids than those stated.

All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons. I have correctly marked all texts or parts of texts that were generated in collaboration with other persons.

In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty.

The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me.”

Date Signature

8. Curriculum vitae

My curriculum vitae does not appear in the electronic version of my paper for reasons of data protection.

9. List of Publication

Chen X.**, Qu Y.**, Sheng L., Liu J., Huang H. and Xu L.* (2014) A simple method suitable to study de novo root organogenesis. Front. Plant Sci. 5: 208 (Co-first author)

10. Acknowledgements

I firstly would like to express my deep and sincere gratitude to my primary advisor Dr. Zsuzsanna Izsvák. I experienced a frustrating beginning of my PhD study. Dr. Zsuzsanna Izsvák took over the supervision and offered me the opportunity to work on this fascinating project. Without her continuous guidance and supports, I could not have gone so far. Many thanks to my PhD committee members, Prof. Dr. Ralf Dechend and Prof. Dr. Thomas Willnow, for their invaluable advice and patience in both my study and the university matters.

I am sincerely grateful to all current colleagues and alumni from the Mobile DNA Research Group at MDC. I would like to thank Dr. Julianna Rugor, Dr. Manvendra Singh, Dr. Amit Pande and Kathrin Radscheit for their generous assistance in data analysis. I highly appreciate the constructive advice in experiments from Dr. Rabia Anwar, Katarina Stevanovic, Aleksandra Kondrashkina, Zhimin Zhou and Dr. Tamas Rasko. A big thank you to Sandra Neuendorf, Beatrice Heuser, Christoph Kempe and Beate Valeske. Without their continuous efforts, we could not have owned the well-organized working environment. I would also like to thank Yong Guo, Felix Lundberg, Bertrand Tangu Teneng, Chigozie Joseph Samuel, Karam Ibrahim for their help and encouragement. We have shared a lot of beautiful memories.

Besides, I would like to take this opportunity to express my heartfelt gratitude to all collaborators and technicians involved in this project. My profound thanks to Dr. Florian Herse, Dr. Sandra Blois and Dr. Gabriela Dveksler for their continuous supports and inspiring discussion in experiments. Many thanks to Christine Molenda from the Advanced Light Microscopy technology platform, Dr. Hans-Peter Rahn and Kirstin Rautenberg from the flow cytometry technology platform, Julianne Anders and Jana Czychi from Müller/Dechend Lab, for their generous patience and technical assistance to this project.

I am also thankful to my roommates Haiyue Liu and Shixian Yan for their warm companionship in past years. Special thanks to Dr. Jichang Wang, Dr. Kun Song, Tao Chen, Dr. Bin Zhang, Dr. Ning Liang for their guidance in my academic training. Lastly, my deepest gratitude goes to my parents, my sister, my brother-in-law and my lovely nephew. Their selfless love has been always the source of my courage and strength to overcome any hardship throughout my life and study abroad.

11. Certificate for statistical consultation

CHARITÉ

CharitéCentrum für Human- und Gesundheitswissenschaften

Charité | Campus Charité Mitte | 10117 Berlin

Institut für Biometrie und klinische Epidemiologie (iBikE)

Direktor: Prof. Dr. Geraldine Rauch

Postantschrift:
Charitéplatz 1 | 10117 Berlin
Besucheranschrift:
Reinhardtstr. 58 | 10117 Berlin

Tel. +49 (0)30 450 562171
geraldine.rauch@charite.de
https://biometrie.charite.de/

Name, Vorname: Qu, Yuliang

Emailadresse: yuliang.qu@charite.de

Matrikelnummer: 226518

PromotionsbetreuerIn: Dr. Zsuzsanna Izsvak,
Prof. Dr. Ralf Dechend, Prof. Dr. Thomas Willnow

Promotionsinstitution / Klinik: Max-Delbrück-Centrum für Molekulare Medizin in der Helmholtz-Gemeinschaft (MDC)

Bescheinigung

Hiermit bescheinige ich, dass Herr *Yuliange Qu* innerhalb der Service Unit Biometrie des Instituts für Biometrie und klinische Epidemiologie (iBikE) bei mir eine statistische Beratung zu einem Promotionsvorhaben wahrgenommen hat. Folgende Beratungstermine wurden wahrgenommen:

- Termin 1: *04.12.2020*

Folgende wesentliche Ratschläge hinsichtlich einer sinnvollen Auswertung und Interpretation der Daten wurden während der Beratung erteilt:

- Varianzanalyse: One-way and Two-way ANOVA
- Korrektur für multiples Testen
- T-Test für unabhängige Stichproben
- Graphische Darstellung

Diese Bescheinigung garantiert nicht die richtige Umsetzung der in der Beratung gemachten Vorschläge, die korrekte Durchführung der empfohlenen statistischen Verfahren und die richtige Darstellung und Interpretation der Ergebnisse. Die Verantwortung hierfür obliegt allein dem Promovierenden. Das Institut für Biometrie und klinische Epidemiologie übernimmt hierfür keine Haftung.

Datum: 07.12.2020 Name des Beraters/ der Beraterin: Pimrapat Gebert

P.Gebert

Unterschrift BeraterIn, Institutsstempel

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN
Institut für Biometrie und
Klinische Epidemiologie
Campus Charité Mitte
Charitéplatz 1 | D-10117 Berlin
Sitz: Reinhardtstr. 58