Expression and Regulation of Pregnancy-Specific Glycoproteins in the mouse

John Michael Williams
BSC

NATIONAL UNIVERSITY OF IRELAND, CORK

DEPARTMENT OF BIOCHEMISTRY

Thesis submitted for the degree of Doctor of Philosophy

June 2013

Supervisor: Dr. Tom Moore
Head of Department/School: Prof. Tom Cotter

Research supported by Science Foundation Ireland

Contents

I, John Michael Williams, certify that this thesis is my own work and I have not obtained a degree in this university or elsewhere on the basis of the work submitted in this thesis.

John Michael Williams

List of Figures

List of Tables

"I hated every minute of training, but I said, 'Don't quit. Suffer now and live the rest of your life as a champion'."

Muhammad Ali

Acknowledgements

I would firstly like to thank my supervisor Dr. Tom Moore for giving me the opportunity to undertake this exciting research project in his lab and whose exceptional guidance, tutoring and endless patience has made this work possible. I would like to thank all the members of the developmental genetics lab, Mel, Ronan, Zhenfei, and Danny, for making work so enjoyable everyday and for the continuous encouragement and help. I would like to give a special thank you to all members of the tea crew, past and present, especially Sarah, Eoin, Conor, Ronan, Megan, Zhenfei, Aoife, Ciara, Catriona, Fi, Sylvia, Lydia, Mike S. and of course Rose.

A big thank you to my Bujinkan brothers; Westy, Macca, Mike and Mark; who introduced me to the path of Budo, and gave me a port of calm in the storm. I have to mention my brother from another mother, Ross, also Bryan BF, Aimee, Johno, Gilly, Stouty and Mr. and Mrs. West whose friendship was unwavering throughout this work, and who listened to my endless waffle. Stephen, even though you are so far away, you have been here for me throughout and for that I am so grateful. Also a big thanks to Sparks and Hans, who were there electronically to brighten everyday. I would like to thank the entire Mullins clan for all their help, love and support throughout this project, Billy, Elma, Marley and Jeff.

Wardy, my brother, you will never be forgotten.

I would like to thank my family, Mum, Dad, and Liam, for their endless support. You are always there to listen, to encourage, and to guide me through the good and difficult times. Your relentless love and encouragement help me realise that I could do whatever I put my mind to. You have been my source of inspiration and endurance through this work, and no amount of thanks will ever suffice. Mom, thanks for all the chats, and Dad, thanks for the genetic compulsion to strive for perfection.

To my dearest Emilie, I couldn't have done this without you. You have been my rock throughout and a source of strength to allow me to complete this. The continual love, support, food and wake up calls, will never be forgotten and I will be forever indebted to you with all my heart.

Abstract

Pregnancy-Specific Glycoproteins (PSG) are the most abundant fetally expressed proteins in the maternal bloodstream at term. This multigene family are immunoglobulin superfamily members and are predominantly expressed in the syncytiotrophoblast of human placenta and in giant cells and spongiotrophoblast of rodent placenta. PSGs are encoded by seventeen genes in the mouse and ten genes in the human. Little is known about the function of this gene family, although they have been implicated in immune modulation and angiogenesis through the induction of cytokines such as IL-10 and TGFβ1 in monocytes, and more recently, have been shown to inhibit the platelet-fibrinogen interaction. I provide new information concerning the evolution of the murine *Psg* genomic locus structure and organisation, through the discovery of a recent gene inversion event of *Psg22* within the major murine *Psg* cluster. In addition to this, I have performed an examination of the expression patterns of individual *Psg* genes in placental and non-placental tissues. This study centres on *Psg22*, which is the most abundant murine *Psg* transcript detected in the first half of pregnancy. A novel alternative splice variant transcript of *Psg22* lacking the protein N1-domain was discovered, and similar to the full length isoform induces *TGFβ1* in macrophage and monocytic cell lines. The identification of a bidirectional antisense long non-coding RNA transcript directly adjacent to *Psg22* and its associated active local chromatin conformation, suggests an interesting epigenetic gene-specific regulatory mechanism that may be responsible for the high level of *Psg22* expression relative to the other *Psg* family members upon trophoblast giant cell differentiation.

Abbreviations

AA - Amino acid

AAS - Antibiotic Antimycotic Solution

Bgp - Billiary glycoprotein

bHLH - Basic helix loop helix

CD9 - Cluster of differentiation 9

cDNA - Complementary deoxyribonucleic acid

CDS - Coding sequence

CEA - Carcinoembryonic antigen

CEACAMs - Carcinoembryonic antigen-related cell adhesion molecules

CMV - Cytomegalovirus

CPE - Core promoter element

CTB - Cytotrophoblast

C-TGC - Canal trophoblast giant cells

DC - Dendritic cells

DNA - Deoxyribonucleic acid

DMEM - Dulbecco's Modified Eagles' medium

E - Embryonic day

EPC - Ectoplacental cone

ER - Endoplasmic reticulum

ES - Embryonic stem

EST - Expressed sequence tag

EtOH - Ethanol

Exe - Extraembryonic ectoderm

FAE - Follicle-associated epithelium

FCM - Fibroblast conditioned medium

Gadph - Glyceraldehyde-3-phosphate dehydrogenase

GIT - Gastrointestinal tract

GlyT - Glycogen trophoblast cells

hGC - Human chorionic gonadotrophin

Hprt - Hypoxanthine-guanine phosphoribosyltransferase

HSPG - Heparan and chondroitin sulfate proteoglycans

ICM - Inner cell mass

IgC - Immunoglobulin constant domain

IgV - Immunoglobulin variable domain

IRE1a - Inositol requiring enzyme-1a

ITIM - Immunoreceptor tyrosine-based inhibition motif

KLF4/6 - Kruppel-like factor 4/6

KO - Knock out

lincRNA - Long intergenic non-coding ribonucleic acid

lncRNA - Long non-coding RNA

MEF - Mouse embryonic fibroblasts

MFC - Maternal foetal conflict

miRNA - Micro ribonucleic acid

MMC - Mitomycin C

mRNA - Messenger ribonucleic acid

NCBI - National Center of Biotechnology Information

ncRNA - Non-coding ribonucleic acid

NEB - New England Biolabs

ORF - Open reading frame

PAC - P1-derived artificial chromosome

PCR - Polymerase chain reaction

PBMC - Peripheral blood mononuclear cells

PP - Peyers' patches

piRNA - Piwi ribonucleic acid

PSG - Pregnancy-specific glycoproteins

P-TGC - Parietal trophoblast giant cells

RA - Retinoic acid

RAR - Retinoic acid receptor

RARE - Retinoic acid response element

RNA - Ribonucleic acid

RGD - Arginine-Glycine-Aspartic acid

RT-PCR - Reverse transcriptase polymerase chain reaction

RXR - Retinoid X receptor

SDS-PAGE - Sodium dodecyl sulfate polyacrylamide gel electrophoresis

shRNA - Short hairpin ribonucleic acid

SP1 - Specificity protein 1

Spa-TGC - Spiral atery trophoblast giant cells

SpT - Spongiotrophoblast

STB - Syncytiotrophoblast

S-TGC - Sinusoidal trophoblast giant cells

TBE - Tris borate EDTA

TE - Trophectoderm

TSC - Trophoblast stem cells

TGC - Trophoblast giant cells

tRNA - Transfer ribonucleic acid

TSS - Transcriptional start site

qRT-PCR - Quantitative Real-time polymerase chain reaction

XBP1 - X-box binding protein 1

UCSC - University of California, Santa Cruz

UNG - Uracil N-Glycosylase

Chapter 1

Introduction

The placenta of eutherian mammals is a remarkable biological structure which originated more than 100 million years ago (mya) [1] and therefore is relatively recent in terms of vertebrate evolution [2]. It is composed of both embryonic and maternally derived cells, and facilitates the complex interactions between the mother and the fetus that are necessary for fetal growth and survival [3]. The mouse (*Mus musculus*), serves as a useful model for studying the development of the haemochorial placenta, as corresponding placental tissues from the human at many stages of gestation are not generally accessible due to legal and ethical constraints. Mouse and human placentas share a discoid shape, hemochorial exchange, analogous cell types and cell layers, and molecular features [4, 5]. There are exceptions to this placental similarity, with idiosyncracies in our fetal membrane development including primary interstitial implantation in a simplex uterus, the lack of yolk sac placentation, and the development of an allantoic stalk rather than an allantoic sac [6]. Murine implantation and trophoblast invasion is shallower and more restricted than in humans [6].

Despite these differences, the mouse is a useful model to investigate the genetic basis of trophoblast development. In addition to anatomical similarities in trophoblast development, these two species demonstrate a large amount of chromosomal synteny and gene orthology, in the developmental regulatory mechanisms of the trophoblast, which is useful for comparative genetic analysis. A major benefit of mice compared to other rodent models lies in the availability

of embryonic stem cells and technologies to produce genetic knockout *(-/-)* and transgenic mice [6]. Null phenotype data generated in the mouse has helped us gain valuable insights into the complexity of the differentiation and regulation of the trophoblast. Gene expression patterns that are conserved in humans should enable the interpretation of the molecular basis of human placental dysfunction [5]. Placental dysfunction and disease can have detrimental effects which contribute to morbidity and mortality in mother and fetus. Preeclampsia, Hydatidaform mole, and spontaneous abortion are a number of pregnancy complications that occur in the human. An understanding of the embryological development of the placenta in a variety of eutherian mammals will facilitate in the treatment and prevention of these common disorders.

The fully developed placenta in humans and rodents is composed of three distinct layers: the outer maternal layer (decidua basilis), which includes decidual cells of the uterus as well as the maternal vasculature that brings blood to/from the implantation site; a middle spongiotrophoblast (SpT) "junctional" region, which attaches the fetal placenta to the uterus and contains fetoplacental (trophoblast) cells that invade the uterine wall and maternal vessels; and an inner labyrinth layer, composed of highly branched villi that is bathed in maternal blood and facilitates efficient nutrient exchange [7]. Each of these layers possess specialised endocrine, paracrine, vascular, immunological or transport functions during gestation [8]. The maternal blood supply passes through this junctional zone via large central 'arterial' sinuses in which the maternal endothelial cells are eroded away and replaced by trophoblast cells. The maternal blood eventually enters into the intricate spaces of the labyrinth where the fetal trophoblastic villi are bathed by maternal blood enabling material exchange between the two blood systems [9]. The fetal trophoblastic villi are composed of outer epithelial layers that are derived from the trophoblast cell lineage and an inner core of stromal cells and blood vessels [7]. It is this invasive form of implantation and direct foetal contact with maternal perfusing blood that is characteristic of haemochorial placentation (Fig:1.1.).

Functionally the placenta is an endocrine organ that produces various

placental hormones and secreted factors that are found in abundance in the maternal bloodstream during pregnancy and are essential for maintaining a suitable environment for pregnancy and fetal development [10]. The fetus is considered to be semi-allotypic in the maternal body; nevertheless, in most cases, immune rejection of the fetus does not occur [11]. At the interface of fetal and maternal tissues, the cells of the innate and adaptive immune systems have been found to produce both Th1- and Th2-type cytokine subclasses. Changes in the cytokine profile is dependent on gestational-age, and in some pregnancy complications, many cytokines have been shown to exert both pro- and anti-inflammatory functions, depending on their binding with their receptors, or intensity, and duration of the stimulation [12, 13]. A multitude of data suggests that maternal immunity is skewed toward the anti-inflammatory Th2 condition during pregnancy, which protects the developing fetus from immune rejection [14]. Initial data demonstrating an immunoregulatory function for the placenta was the discovery of high expression of HLA-G in human trophoblasts [15]. It is these trophoblast cells that secrete placenta-specific hormones that are responsible for the immunomodulatation of the maternal physiology and also fulfill a variety of structural and functional roles in the haemochorial placenta. The cells of the trophoblast lineage constitute the epithelial compartment of the placenta, and the establishment and maintenance of pregnancy is dependent on the precise development of these cells [16].

1.1 The Trophoblast

1.1.1 Development of the trophoblast

Once the embryo is anchored within the uterine wall, the next major event is the formation of the extraembryonic lineages, a necessary prelude to assembly of the maternal-fetal interface [17]. Only recently, considerable insights have been gained into how the trophoblast lineage differentiates at the blastocyst stage due to the generation over 100 mutant mouse lines that manifest defects in placental development. In addition, the derivation of murine trophoblast stem cells (TSC)

has provided a powerful resource for understanding the molecular mechanisms governing TSC maintenance and differentiation [16]. In all mammals, the trophoblast cell lineage is specified before implantation. Implantation involves a succession of genetic and cellular signals [18], that implement a reciprocal interaction mediating apposition and adhesion between trophectoderm (TE) in the blastocyst and uterine epithelium, followed by trophoblast invasion [19]. In mice at embryonic day (E) 3.5, placental development begins when this lineage appears in the blastocyst as the TE, a sphere of epithelial cells surrounding the inner cell mass (ICM) and the blastocoel at around the 32-cell stage of development [20, 21, 22]. The appearance of a progenitor population of TSC represents the initial differentiation event of embryogenesis [23]. At this stage they have not fully committed to a definite cell fate, as evidence has been found that the outside cells of the late morula can produce ICM derivatives [24], and inside cells can make trophoblast tissue [25]. It is not until blastocyst formation that the TE and ICM lineages are irreversibly determined [26].

The TE layer of the preimplantation embryo is the precursor to all trophoblast cell subtypes (Fig:1.1. and Fig:1.2.). However, the entirety of the TE layer does not contribute equally to the various trophoblast subtypes. Upon implantation, the TE of the blastocyst not in contact with the ICM, designated the mural TE, differentiates to form post-mitotic primary trophoblast giant cells (TGC) that migrate into the antimesometrial portion of the implantation chamber and surround the future parietal yolk sac. In contrast, the TE directly overlying the ICM, known as the polar TE, retains its capacity to proliferate and expands to form the extraembryonic ectoderm (ExE) and ectoplacental cone (EPC) [27]. The chorion, which is a tightly packed layer of TSC, is in contact with the base of the EPC. TSC are defined as pluripotent cells whose differentiated derivatives are restricted to the trophoblast lineages. The restricted potential of TSC to exclusively contribute to trophoblast-derived cell types has been demonstrated in chimeras *in vivo* where they can they can give rise to all trophoblast elements of the mouse placenta, but they are unable to contribute to the embryonic germ layers giving rise to the tissues of the fetus [21]. Stem cell potential is maintained in trophoblast cells of the ExE post-implantation. This is reflected by the ability to

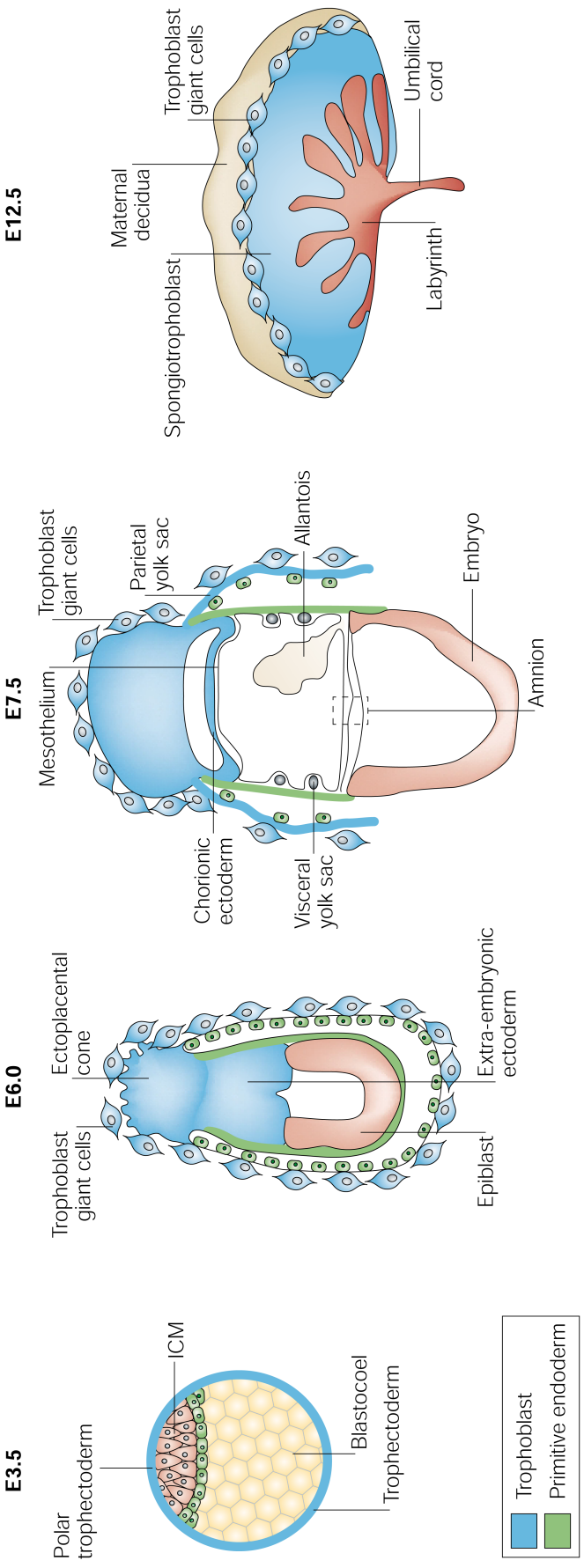


Figure 1.1: Placental development in the mouse. Early development of the mouse embryo from embryonic day (E) 3.5 – E12.5, showing the origins of the extra-embryonic lineages and the components of the placenta. ICM, inner cell mass. (modified from [9]).

derive morphologically and functionally indistinguishable TSC lines from blastocyst stage embryos as well as from the ExE and its derivatives in the chorion until E8.5 [21].

The TSC of the chorion layer develops from the ExE, and later some of these TSC will differentiate towards a labyrinth fate. During this development, the labyrinth is structurally supported by the SpT which is derived from the EPC. The vasculature of the placenta is derived from the extraembryonic mesoderm of the allantois that extends from the posterior end of the embryo at E8.0. The junction of the allantois and the chorion joins together at E8.5, in a process called chorioallantoic fusion, even though no physical cell fusion occurs [9]. After chorioallantoic fusion takes place, folds begin to form in the chorion which develop into the villi, creating a space into which the fetal blood vessels grow from the allantois and this becomes the fetal component of the placental vasculature [31]. The labyrinthine villi become larger and more extensively branched until birth (E18.5–19.5). Around E11.5, the labyrinth and junctional zones are indistinguishable and consists of strands of SpT and TGC, separated by maternal blood sinuses. Glycogen trophoblast cells (GlyT) appear at E12.5 and at this stage the labyrinth and junctional zones are distinguishable [32]. These GlyT differentiate within the SpT layer, and form a dense layer of non-syncytial cells between the labyrinth and the outer giant cells, which consequently diffusely invade the uterine wall and corresponds to the column cytotrophoblast (CTB) of the human placenta [33, 9]. Differentiating trophoblast cells acquire specialized functions that are essential for the establishment and maintenance of pregnancy including: invasion, nutrient and waste transport, metabolism, protection from the maternal immune system, and production of hormones and cytokines that likely contribute to all of these functions. Progression along the trophoblast lineage is dictated by the activation of sets of genes characteristic of the specific differentiated trophoblast phenotype [34].

1.1.2 Trophoblast Giant cells

In rodents, the most invasive of the placental cells are the TGC, so named because of their unusually large size which is related to the fact that they are extensively

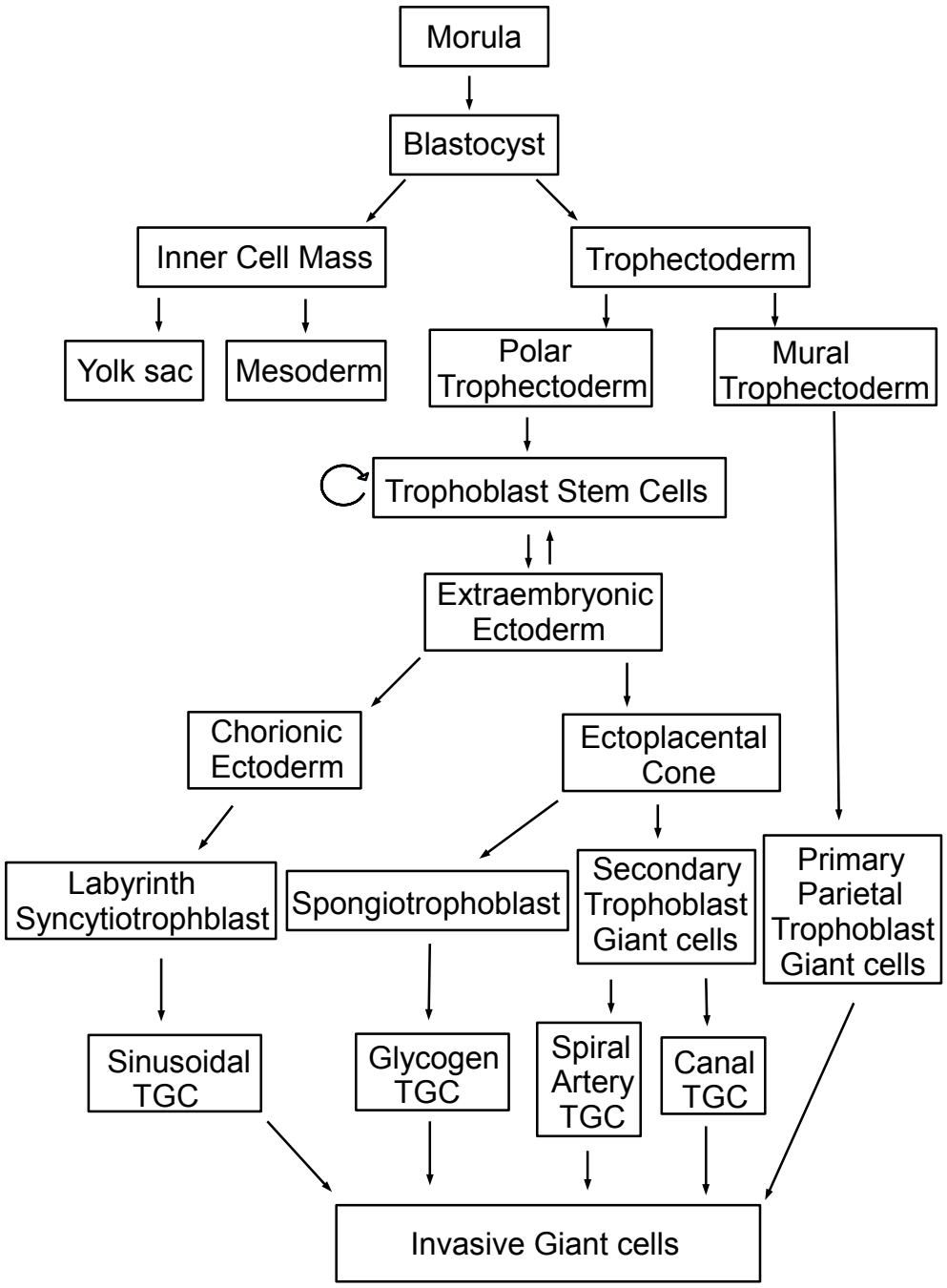


Figure 1.2: Trophoblast lineage and origins of TGC subtypes. (modified from [16, 28, 29, 30]).

polyploid and terminally differentiated cell types. In rodents, TGC are the first terminally differentiated subtype of cells to be derived from the trophoblast cell lineage [35]. Proliferative trophoblasts differentiate into TGC as they exit the cell cycle and enter a process of endoreduplication, an unusual cell cycle with successive rounds of deoxyribonucleic acid (DNA) synthesis in the absence of intervening mitoses [36, 37]. TGC in the rodent placenta form the outermost layer of the extraembryonic compartment. This layer is responsible for establishing direct contact with maternal cells facilitating in embryo implantation, conceptus invasion, and provides a number of pregnancy-specific cytokine hormones [8, 38].

The mural trophectoderm, trophectoderm cells which are not in contact with the ICM at the time of implantation (E4.5), stop dividing and differentiate to form a limited number of TGC which line the implantation chamber, anastomosing to form a diffuse network of blood sinuses for the early transport and exchange of nutrients and endocrine signals [39]. These cells are analogous to human extravillous cytotrophoblast cells [9]. The trophectoderm immediately overlying the ICM, the polar trophectoderm, continues to proliferate and gives rise to all the remaining trophoblast cell types of the placenta [20], including SpT, glycogen trophoblast cells, several labyrinth trophoblast cell types, and a later influx of TGCs (called 'secondary' to distinguish them from the initial 'primary' group) [16, 40]. Four TGC subtypes have been identified in the placenta each of which posses specialised functions and are listed in Table 1.1, these TGC subtypes include parietal TGC (P-TGC), that line the implantation site and are in direct contact with decidual and immune cells in the uterus, spiral artery-associated TGCs (SpA-TGC), maternal blood canal-associated TGCs (C-TGC), and sinusoidal TGC (S-TGC) that are within the sinusoidal blood spaces of the labyrinth.

These TGC subtypes share common characteristics like their large size, invasive, phagocytic and secretory nature [39, 41]. Even so, the four subtypes of TGCs can be distinguished by their anatomical location and gene expression [16]. These four distinct TGC subtypes are derived from different TE lineages origins at different periods during placentogenesis [35]. The gene expression markers that correspond

to these four TGC subgroups are shown in Table 1.1, some of which I utilised in the characterisation of TGC populations that were present in a culture of differentiated TSC. P-TGCs arise directly from approximately 60 mural trophectoderm cells in the blastocyst in a process called primary TGC differentiation, although the several hundred P-TGCs that are present by mid-gestation, and all of the other TGC subtypes, arise from the polar trophectoderm through so-called secondary TGC differentiation. Both P-TGC and C-TGC have mixed developmental origins. In contrast, all of the Spa-TGC originate from *Tpbpa* positive cells, whereas all of the S-TGC originate from *Tpbpa* negative precursors [29]. The locations of these four types of TGC in the mature murine placenta are shown (Fig:1.3.).

Table 1.1: TGC subtypes in the mature placenta (modified from [29]).

Subtype	**Location**	**Temporal appearance**	**Marker genes**	**Suggested function**
SpA-TGC	Lining maternal spiral arteries bringing blood into placenta	E10.5	*Plf*	Regulate maternal spiral artery remodeling and blood flow into the placenta
P-TGC	Lining implantation site and outer layer of parietal yolk sac	E7.5	*Pl1, Pl2, Plf*	Facilitate implantation and initial maternal vascular connections, regulate decidual cell differentiation, and maternal physiology
C-TGC	Lining canals that bring maternal blood to base of labyrinth	E10.5	*Plf, Pl2*	Regulate maternal vasculature remodeling and maternal physiology
S-TGC	Within maternal blood sinusoids of the labyrinth layer	E10.5	*Ctsq, Pl2*	Modulation of hormone and growth factor activity before they enter fetal and/or maternal circulation, regulate maternal physiology

In the mouse, two different phases of trophoblast invasion can be distinguished, these are either endo- or perivascular, as invasive trophoblast cells are strictly associated with maternal arteries where they displace endothelial cells or are located within one or a few cell layers underneath the vascular endothelium [33, 42, 43, 44]. Endovascular TGCs, invade great distances into the maternal spiral arteries to replace endothelial cells and express *Plf* but not *Pl1*. Endovascular TGCs more proximal to the placenta express both genes [45]. TGCs produce *Pl-1* starting soon after implantation until mid-gestation and subsequently *Pl-2* from mid-gestation until term [46]. The morphology of endovascular TGCs is also clearly different to that

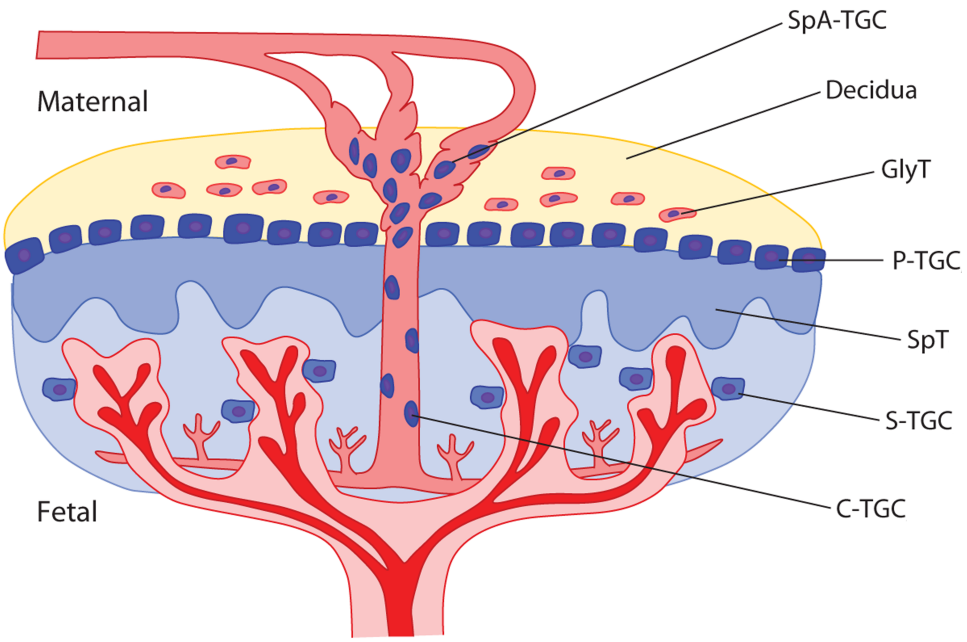


Figure 1.3: Locations of TGC subtypes in mature placenta. SpA-TGC, spiral artery trophoblast giant cell; GlyT, glycogen trophoblast; P-TGC, parietal trophoblast giant cell; SpT, spongiotrophoblast; S-TGC, sinusoidal trophoblast giant cell; C-TGC, canal trophoblast giant cell. (modified from [2, 7]).

of interstitial TGCs as they are much smaller and more spindle-like [45].

Only after gestational day (E) 14.5, a different, 'interstitial' type of trophoblast invasion is observed where cytokeratin-positive trophoblast cells are broadly penetrating into the decidual stroma and are not obviously associated with maternal blood vessels. Morphological characteristics such as a vacuolated-appearing cytoplasm, a positive PAS stain and expression of the SpT marker gene *Tpbpa* identify these cells as glycogen cells [44]. SpT cells comprise the middle layer of the placenta sandwiched between the outer secondary TGCs and the inner labyrinth layer (Fig:1.3.). The function of the SpT layer (or Junctional zone) is poorly understood. However, it could act as structural support for the developing villous structures of the labyrinth and is also known to express several unique genes. Precursors for SpT cells reside within the EPC. However, observations from several mouse mutants suggest that SpT and TGC can arise from a common EPC precursor [47, 48].

TGCs have diverse functions that are crucial for implantation and subsequent placental function. The mural trophectoderm-derived TGCs mediate attachment of

Figure 1.4: Paracrine and endocrine function of TGC in mature mouse placenta. TGCs have diverse functions that are crucial for implantation and subsequent placental function. LacZ stained GC-GFP cell (modified from [29]).

blastocyst to the uterine epithelium, induce uterine decidulization, invade into the uterine stroma, and anastomose to form the yolk sac placenta for early exchange of nutrients and endocrine signals between mother and fetus. After implantation, TGCs produce hormones and cytokines for maintenance of the feto-maternal interface and regulation of maternal adaptations to pregnancy [29]. The four TGC subtypes have a variety of functions, and these TGC exert their functions through a multitude of paracrine and endocrine mechanisms (Table 1.1.). These paracrine and endocrine effects of TGC, including the responsible signalling molecules are shown (Fig:1.4.). This figure demonstrates the diverse roles TGC play in the initiation and maintanence of pregnancy, from implantation and vascular remodelling to modulating maternal immune physiology and adaptive behaviour. Some of the TGC functions may be mediated by their ability to produce Pregnancy-specific glycoproteins (PSG), as suggested by the prominent association of at least one *PSG* with the endothelial lining of vascular spaces surrounding the implantation site from E8.5 to E11.5 [49]. TGC function depends on successful differentiation from TSC. In order for trophoblast proliferation and differentiation to occur properly, a specific microenvironment must exist to support the maintenance of the trophoblast stem cell population [50], which relies on a complex regulatory signalling mechanisms, which are discussed in the next section.

1.1.3 Trophoblast regulatory pathways

ES cells predominantly contribute to the embryo proper and TSC only contribute to the various trophoblast cell types of the placenta. Along with studies of mouse mutants these stem cell lines have allowed us to begin to elucidate the transcriptional networks that define the two earliest cell populations and orchestrate lineage-specific transcriptional programmes in all their progenitor cells [51]. Like other stem cells, TSC cells have the ability to self-renew or to differentiate into more specialized, lineage-specific cell types, depending on reception of appropriate signals [52]. Maintenance of trophoblast proliferation and self-renewal is dependent on signals from the ICM. Indeed, ICM cells inserted into an empty sphere of trophectoderm can induce

secondary sites of proliferation [27]. Pluripotent trophoblast stem cells reside within the extraembryonic ectoderm and later the chorionic ectoderm [21, 53], and provide the EPC with progenitors which give rise to the SpT and secondary TGC. Maintenance of trophoblast proliferation later in gestation is also dependent on close proximity with the embryonic-derived epiblast as isolated EPC or EXE transplanted ectopically [54]. Signalling pathways establish the transcriptional circuitry that underpins TE identity and how the core trophoblast transcription factors coordinate lineage commitment, maintenance of the stem cell niche and eventual differentiation into placental cell types [55]. Due to the changing nature of embryonic and ExE development, it is probable that this specialized niche is temporary and exists only for 3–4 days during post-implantation development [22, 53].

The maintenance of TSC in the early embryo is dependent on FGF signaling involving the ligand Fgf4, which is a paracrine factor produced by the ICM/epiblast and signals through MAPK and controls trophoblast proliferation [56]. *Fgf4* is expressed in early embryos, becoming restricted to the ICM of the blastocyst and later to the epiblast of the early post-implantation embryo [57, 58]. TSC maintenance is also dependant on the FGFR2 receptor, which is expressed in trophoblastic tissues, including the ExE and chorion [9, 59, 47]. As mentioned previously, when cultured in the presence of Fgf4, mouse TSC exhibit sustained undifferentiated proliferation, without significant expression of the phenotypic markers of placental trophoblasts, such as *Pl-1/Pl-2*, or placental prolactin-related proteins. Removal of Fgf4 results in the arrest of cell proliferation, rapid TGC formation and onset of hormone gene transcription [21]. *Fgf4* expression is induced by the TGFβ-related protein Nodal. Nodal, along with Fgf4, acts directly on adjacent ExE that maintains a microenvironment that inhibits premature TSC differentiation [50]. Nodal plays an important role in trophoblast differentiation, as conceptuses that possess a hypomorphic mutation in *Nodal*, result in an expansion of the TGC and SpT layers, and a decrease in labyrinthine development [60]. The addition of Fgf4 alone can inhibit the induction of *Mash2* but cannot maintain expression of *Cdx2*, *Eomes*, and *Err2*. Conversely, addition of Nodal or Activin alone cannot inhibit *Mash2*

expression in ExE but, in combination with Fgf4, can maintain *Cdx2, Eomes*, and *Err2* expression [50]. Activin or TGFβ1 can also replace MEF conditioned medium for the maintenance of TSC proliferation *in vitro* suggesting that constitutive FGF signaling in TSC selectively inhibits the ability of TGFβ1 to repress *c-myc* expression, a central component of the TGFβ1 cytostatic transcriptional response [61]. Fgf4 produced in the embryonic ectoderm, signals through Fgfr2 to maintain the expression *Cdx2, Eomes*, and *Err2* and suppression of *Mash2* expression in the ExE. Nodal produced in the epiblast maintains *Fgf4* expression and cooperates with Fgf4 to maintain TSC marker expression in EXE [16]. *Mash2* is required for the maintenance of TSC and is essential for maintaining SpT cells at the expense of TGC differentiation, as in its absence, the SpT layer is lost and an excess of TGC form [21, 62]. *Mash2* overexpression prevents TGC differentiation and the suppression of *Mash2* function, required to allow TGC differentiation, may occur *in vivo* by loss of its E-factor partner due to loss of its expression and/or competition from *Hand1* [63].

Once TE and ICM lineages are delineated, it is clear that the POU domain transcription factor, *Oct4*, has an important role in ICM fate determination [16]. Regulatory sequences of the *Oct4* gene are hypermethylated and associated with a closed chromatin structure in TSC, whereas these regions are hypomethylated with an open chromatin structure (acetylated histones) in ES cells, resulting in differential gene expression [64]. *Oct4* has been shown to directly repress the transcription of several trophoblast-specific genes [65, 66, 67]. *Sox2* has a similar function in repressing the trophoblast cell fate as that observed for *Oct4* [68] and works together with *Oct4* to regulate down-stream targets expressed in the ICM [69]. It has been shown that less than a 2-fold increase of Sox2 protein levels in ESC is sufficient to down-regulate *Nanog* and drive trophoblast, mesodermal and ectodermal differentiation [70]. *Elf5* has an important role as its epigenetic regulation by DNA methylation positions it as a gatekeeper of cell lineage fate between the trophoblast and embryonic compartments. *Elf5* expression is found from the late blastocyst stage onwards in the EXE where it maintains the expression of *Cdx2* and *Eomes* [71, 72]. Consistent with its expression in trophoblast cells, the *Elf5* promoter is unmethylated in TSC, but methylated in

ES cells where *Elf5* is not expressed [72]. This differential epigenetic modification of *Elf5* establishes a stable cell lineage barrier between the embryonic and trophoblast compartments as it restricts the positive transcriptional feedback loop between *Cdx2*, *Eomes* and *Elf5* to the trophoblast lineage [55].

TGC differentiation is determined through a similarly complex transcription factor signalling regulation as is with TSC self renewal. Some of the transcription factors involved in TGC differentiation are shown (Table 1.2.). TGC differentiation depends upon the coordinated activity of a family of transcription factors, most notably basic helix–loop–helix transcription factors (bHLH) [16]. Members of the bHLH family are thought to function as heterodimers, typically between the cell subtype-specific factors and the widely expressed E proteins, such as E12/E47 (which are products of the E2A gene) [73]. While *Mash2* restricts differentiation of TGC, other bHLH genes have the opposite effect. *Hand1* promotes the formation of TGC. The *Hand1* transcription factor is required for TGC differentiation as *Hand1* deficient conceptuses die between E7.5 and E8.5 due to a block in TGC formation, placental defects and noticibly smaller EPC [63, 74]. It has been suggested that *Hand1* could antagonize *Mash2* function by competing for E-factor binding *in vitro* [63]. Other bHLH factors are implicated in trophoblast development based on specific expression patterns. *Hand1*, *Stra13* and *Gcm1* transcription factors override FGF signaling to promote terminal differentiation of TSC [75]. *Stra13* mRNA expression has been suggested in TGC in mice, though not well documented [76] and the bHLH antagonist *I-mfa* promotes TGC differentiation as shown by targeted deletion of *I-mfa* in a C57Bl/6 background which resulted in embryonic lethality around E10.5, associated with a placental defect and a markedly reduced number of TGC. Overexpression of *I-mfa* in rat trophoblast (Rcho-1) stem cells induced differentiation into TGC [77], possibly by inhibiting the function of *Mash2*. Thus, an opposing network of bHLH transcription factors and bHLH interacting proteins regulate TGC differentiation.

As already stated these bHLH factors work alongside a number of other transcription factors that induce TGC formation. These include one of the best-studied determinants of trophoblast cell fate, which is the caudal-type homeobox gene *Cdx2*

Table 1.2: Transcription factors implicated in TGC differentiation

Transcription factor	Relevance to TGCs	References
AP-2	TGC differentiation	[8, 78]
Hand1	TGC terminal differentiation	[8, 63, 75]
Tead4	Trophoblast speciation	[79, 8]
Cdx2	Regulates TE differentiation	[8, 28, 80]
Gata2/3	Regulates TGC differentiation	[29, 81, 82]
Stat3	TGC terminal differentiation	[83, 84]
Ik3	Trophoblast invasion	[84]
RxR	TGC terminal differentiation	[8, 32, 35]
Klf4	Promotes TGC differentiation	[85, 86]
FoxD3	Inhibits TGC differentiation	[28, 87, 88, 89]
NeuroD1	Human CTB differentiation	[84, 90]
Gcm1	TGC terminal differentiation	[75, 91, 92]

[80]. *Cdx2* is required to restrict expression of the pluripotency factors *Oct4* and *Nanog* to the ICM and, even though *Cdx2* null embryos form blastocysts, they fail to maintain trophectoderm cell identity, instead forming a ball of *Oct4* expressing cells incapable of hatching from the zona pellucida [26], demonstrating that *Cdx2* is crucial to maintain a functional TE cell population and is a critical determinant of trophectoderm identity [93]. *Cdx2* is the earliest known factor to have a role in trophoblast lineage development, although the molecular targets mediating its role in trophectoderm identity are still unknown [16]. *Cdx2* is a common marker used to distinguish between TE and ICM cells in the mouse [22, 94]. Interestingly *Cdx2* expression is lost as TSC differentiate to the TGC cell fate.

Even though *Cdx2* and *Oct4* play an essential role in inhibitory feedback signalling in TE lineage differentiation, *Tead4*, is the transcription factor that exerts most influence in TE lineage specification. *Tead4* is required for specification and development of the TE lineage, which includes modulation of *Cdx2* expression [95, 79]. *Tead4* triggers, directly or indirectly, the expression of *Cdx2* and other transcription factors. Once specified, a positive feedback loop involving *Cdx2, Eomes, Tcfap2c,* and *Elf5* reinforces trophoblast identity. In addition to supporting this network, *Gata3, Elf5* and *Ets2* subsequently act to drive further differentiation of the lineage into different placental cell types [55]. The product of the T-box gene *Eomes* is the earliest-acting transcription factor known to be required for immediate post-implantation lineage commitment steps, as mice lacking *Eomes* gene expression fail to exhibit a proper

TE to trophoblast transition. While they do implant, they arrest at a blastocyst-like stage of development [96]. Eomes acts later in TE differentiation than *Tead4* and *Cdx2* by enhancing *Cdx2* expression and promoting the expansion of the EXE [80, 96]. *Eomes* can be activated directly by *Elf5* and *Tcfap2c*, and directly or indirectly by *Cdx2* [79]. The AP-2 family members are also involved in the regulation of human villous cytotrophoblast differentiation. Two of the isoforms, AP-2 α and AP-2 β, are expressed in the human placenta. *AP-2* binding sites are present on the promoters of other genes in the placenta that affect placental function, such as TGFβ1, vascular endothelial growth factor (VEGF), matrix metalloproteinases, tissue inhibitor of metalloproteinases, and the estrogen receptor [97]. *AP-2* γ (also termed *Tcfap2c*) is important in TGC differentiation since it activates the human prolactin promoter [98]. Trophoblast fate induced by *Cdx2* does not require *Tcfap2c*. However, activation of *Elf5* is only achieved in the presence of both factors [99]. *Tcfap2c* cooperates with *Cdx2* to maintain trophectoderm formation, suggesting that *Tcfap2c* and *Cdx2* act in alternate pathways and are both required for the full establishment of TS cell identity [55].

Gata2 and *Gata3* transcription factors have been implicated in the regulation of trophoblast-specific genes [82]. Ray *et al*, 2009, demonstrated that *Gata2* expression was induced during TGC differentiation and hypothesised that *Gata3* directly represses *Gata2* in undifferentiated trophoblast cells, and a switch in chromatin occupancy between *Gata3* and *Gata2* (*Gata3/Gata2* switch) induces transcription during trophoblast differentiation, which regulates a variety other trophoblast-specific genes [81]. *Gata3*-mediated trophoblast fate does not depend on *Cdx2* expression. Considering both these genes are regulated by *Tead4*, they appear to operate semi-independently, specifying trophoblast fate through many different pathways and targets [100]. There are a multitude of other transcription factors whose involvement have been implicated in TGC differentiation, such as *Ets2*, *Ik3*, *Stat3*, *Klf4*, *NeuroD1*. In addition to cell intrinsic factors, extrinsic factors also influence TGC formation. Retinoic acid, for example, can promote TGC formation both *in vitro* and *in vivo* [101], similar to the effects of overexpression of the retinoic acid responsive gene *Stra13* [75]. These complex regulatory pathways and the genes that convey these signals have

been reviewed extensively by [102, 75, 16, 103, 51, 8, 55, 22]. Detailed studies of the cellular and molecular mechanisms governing TSC and TGC formation should give insights into human gestational diseases that are associated with human extravillous cytotrophoblast cells [104].

1.2 CEA superfamily: CEACAMs and PSGs

1.2.1 Ceacams

The carcinoembryonic antigen (CEA) family, which includes two mutligene subfamilies; the CEA-related cell adhesion molecules (CEACAMs) and the Pregnancy-specific glycoproteins (PSGs), are members of the immunoglobulin superfamily [105]. The CEACAM/PSG primordial gene is thought to be common to both primates and rodents, but subsequent gene duplications have arisen independently in both organisms [106]. Gene duplication and conversion is known to be critical to the evolution of gene families [107, 108]. Kammerer *et al*, state that gene families are formed through gene duplications produced by environmental adaptation, which provide new raw genetic material that can be modified by natural selection, without losing the function of the original gene [109]. Haig *et al*, in 1993, hypothesised that antagonism between maternal and fetal genes in the placenta that regulate maternal resource allocation and investment in pregnancy, represents an environment of evolutionary conflict and therefore drives the evolution of these genes [110]. It has been shown that the CEA family, with a subset of other placentally associated genes experience positive selection and rapid evolution based on their pattern of sequence divergence [111].

The CEA subgroup members are cell membrane associated and are expressed in normal and cancerous tissues with notably CEA showing a selective epithelial expression [112]. The nomenclature of the CEACAM family has changed; for example, the original biliary glycoprotein (Bgp), later classified as the CD66a antigen, has now become CEACAM1 (for current and historic nomenclature of the CEACAM family

see [113]). Two different groups identified and characterized CEA complementary deoxyribonucleic acid (cDNA) in 1987 [114, 115]. There are 12 human and 15 mouse CEACAM proteins (Fig:1.5.), CEACAM family members are characterized by a membrane distal IgV-related, N-domain and variable number of IgC2-related domains. A 20 amino acid (aa) leader-like peptide is encoded by the first exon of all CEACAM members, and the second exon codes for the first N-terminal domain (or N-domain) of the mature protein. This N-domain resembles the immunoglobulin variable portion of an Ig molecule, whereas the other exons individually code for the Ig-constant-like domains [116]. CEACAM domain structure shows more variability between family members than the PSGs.

These proteins are linked to the membrane by either a glycophosphatidyl anchor, or by a transmembrane anchor. The cytoplasmic domain can harbour immunoreceptor tyrosine-based inhibition motifs (ITIM), immunoreceptor tyrosine-based switch motifs (ITSM) or ITAM [117]. CEACAMs have a high level of alternatively spliced transcripts. CEACAM1 is the most widely expressed member of the CEA gene family and CEACAM1 is expressed on a number of different cell types including epithelial, endothelial and in a variety of immune cells including B cells, T cells, NK cells, dendritic cells (DC), macrophages and granulocytes [118, 119, 120, 121] and mediates cell–cell adhesion [122]. These interactions are predominantly mediated by the IgV-like N-terminal domain and appear to involve one of the two β-sheets (the CFG-face) of the Ig-fold [123]. Tan *et al*, revealed that based on crystal structure, the degree of variability in sequence of the N-terminal domain for all available mammalian CEA molecules shows that, within the CEA family, most of the variation occurs on the CFG faces of these molecules [124].

Structural and functional analyses show that homotypic and heterotypic adhesion is the most prominent function of these extracellular domains, whereas the cytoplasmic domain is involved in cell growth inhibition and signal transduction [125]. CEACAM1 was found to be one of the pivotal receptors promoting the signalling of immune cells, which is supported by its prominent homophilic adhesion function. CEACAM1, thus seems to be a receptor targeted by pathogens to infect

cells and simultaneously disrupt well coordinated immune responses. The function of CEACAM family members as pathogens receptors, as well as their support of a successful outcome of pregnancy, suggests that pathogen-mediated and fetal-maternal conflict-induced selection are potential key drivers of CEA family evolution [109]. CEACAM5 has been found to possess high expression in adenocarcinomas and other cancers, while CEACAM1 expression is down regulated in many tumors and it has been shown to have a function in tumor-suppression [126]. CEACAMs have diverse roles, with functions in shaping the architecture of epithelia, modulation of T cells and tumor suppression [113].

1.2.2 Pregnancy-Specific Glycoproteins

PSGs are members of a rapidly evolving multigene family [111], and are the most abundant fetal protein in the maternal blood at term in pregnancy [127]. The maternal serum level of these proteins increases with gestation progression and reaches up to 200-400 μg/ml at term, far exceeding the concentration of human chorionic gonadotropin (hCG) and α-fetoprotein [127, 128]. PSGs are produced by cells of the trophoblast lineage; syncytiotrophoblast in higher primates and SpT or TGC in rodents [129, 130, 131]. Just like the CEACAMs, murine *Psgs* are clustered on chromosome 7 in a region syntenic to human chromosome 19q13.2 [132, 133]. PSGs, like many placental hormones, are found in multi-gene families in all species in which they are detected [134]. There are 11 human *PSGs*, 17 murine *Psgs*, and 8 rat *PSGs*. PSGs have been found in a multitude of species that possess haemochorial placentation, like the bat [109]. More recently evidence for their expression in horse has been gathered by searching for evidence of secreted CEACAM related genes in the genome and identifying two related Expressed Sequence Tags (ESTs) from horse trophoblast cDNA libraries [109]. Their expression is localised to the highly invasive portion of the placenta and low *PSG* levels in maternal serum has been correlated with poor pregnancy outcomes, particularly in diseases characterised by placental insufficiency. This indicates that they may play a fundamental role in the formation and maintenance of the maternofetal unit [129, 135, 136, 137, 138, 139, 140, 141, 142,

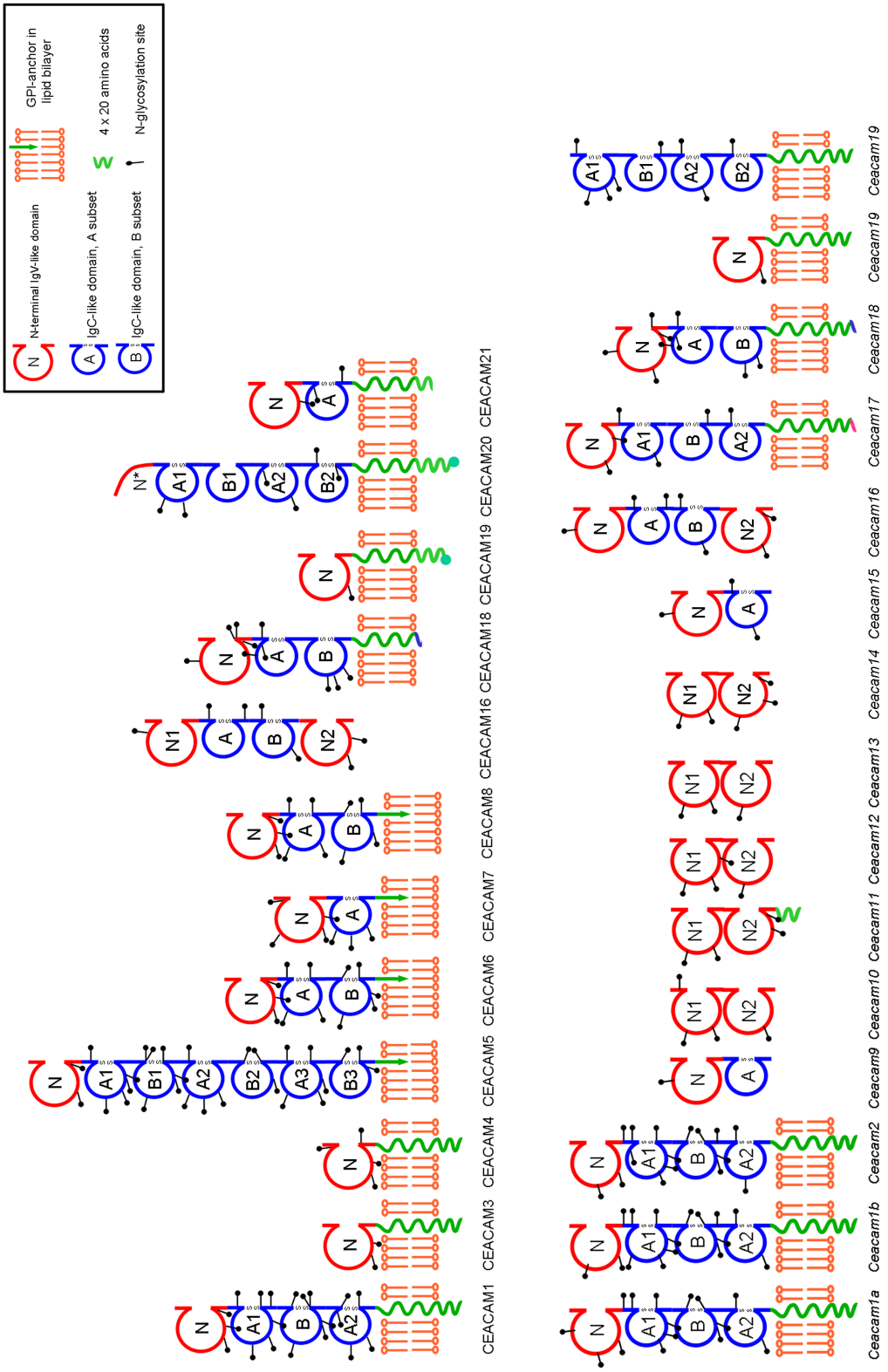


Figure 1.5: Schematic representation of human and mouse Ceacam family domain organisation. (modified from http://www.carcinoembryonic-antigen.de/index.html).

143]. Their importance to the maintenance of pregnancy is further underlined by the observation that the application of anti-PSG antibodies or vaccination with PSG induces abortion in mice and monkeys, respectively, and reduces the fertility of non-pregnant monkeys [144, 145]. However these are old papers and the reliability of the antibodies used is questionable.

In terms of domain architecture and arrangement, *PSGs* are very similar to the CEA-related Cell Adheasion Molecules (CEACAMs) possessing a series of Ig domains in varying numbers and also being highly glycosylated. Comparison of the domain organization of rodent and human *PSGs* reveals a remarkable evolutionary divergence between species. The Ig domain structure of the human and rodent *PSGs* differs between species: Human *PSGs* contain one V-like Ig domain (N), C2-like Ig domains (A and B) and relatively hydrophilic tails (C), with domain arrangements classified as type I (N-A1-A2-B2-C), type IIa (N-A1-B2-C), type IIb (N-A2-B2-C), type III (N-B2-C) and type IV (A1-B2-C). In contrast, rodent *PSGs* typically have three or more N domains followed by a single A domain. All rat *PSGs*, with the exception of *PSG36* (N1-N2-N3-N4-N5-A), are of the N1-N2-N3-A domain arrangement [146]. In contrast, the murine *Psg* family has 14 members which encode a common structure of three Ig variable (IgV)-like domains (N-domains) and a single Ig constant (IgC)-like domain (A-domain) (N1-N2-N3-A) arrangement, and *Psg24*, *Psg30* and *Psg31* which have an expanded structure created by the duplication of (IgV)-like domains. *Psg24* with (N1-N2-N3-N4-N5-A), *Psg30* with (N1-N2-N3-N4-N5-N6-N7-A) and *Psg31*, which possesses a unique duplicated N1 domain, and has a (N1-N1-N2-N3-N4-N5-N6-N7-A) domain arrangement. At the amino acid level the N1 domains of rodents and the N-terminal domain of human PSGs have high similarity. The relatively smaller number of *PSG* genes identified in the rat (compared to the mouse) and the higher level of gene homogenisation implied by split decomposition analysis suggests that the rat *PSG* gene family has not expanded or diversified as extensively as the mouse [146]. The rodent and primate PSGs and CEACAMs common ancestor was most likely similar to *CEACAM1*, which is the only CEA family member with homologous gene structure in the human, rat and mouse that encodes all types of

extracellular domains present in CEACAM and PSG proteins [146, 147, 148, 106].

One of the most striking differences between the CEACAMs and the PSGs across species is the lack of a C-terminal membrane targeting component in PSGs. In CEACAMs a hydrophobic transmembrane sequence or GPI anchoring leads them to attach to the cell surface as opposed to PSG, which appear to be secreted because most of them lack hydrophobic C-terminal domains suitable for membrane anchorage [106]. The schematic arrangement of the immunoglobulin and immunoglobulin-like domains of the 17 mouse, 8 rat and 10 human proteins are shown (Fig:1.6.), which demonstrates the high level of structural conservation based on protein sequence similarity between members of the different PSG families across species. Much current work has focused on human *PSGs* due to their possible relevance to disorders of pregnancy. Nevertheless, the investigation and analysis of rodent *PSG* is significant due to the extensive conservation of expression of these genes in trophoblast, the independent gene family expansions of these genes in mammalian lineages that possess haemochorial placentation, and the implicated conservation of immmune functions during pregnancy [146].

Due to the high levels of conservation of expression and structure, one may assume that human and rodent *PSGs* share a common function. As the N1 domain is the only domain of identical type (IgV-like) and position (first domain) common to rodent and human *PSGs*, it probably plays a major role in determining a conserved function. The tripeptide sequence Arg/Gly/Asp (RGD) found on the CFG face in the N1 domain human *PSGs* is known to be responsible for the interaction of some extracellular matrix proteins with cell surface receptors of the integrin family [149, 150, 151]. *PSG1* is the only human *PSG* that contains a KGD tripeptide motif, rather than RGD tripeptide motif present in the N-domain of the protein. Unlike most primate PSG N domains, rodent *PSG* N1 domains do not possess an RGD tri-peptide motif, but do contain RGD-like motif sequences, which are not found to be conserved in the N2 and N3 domains of rodents. In rodents, and especially in mice, the RGD motif is replaced by a motif which contains a highly conserved Gly residue flanked by a positively and/or a negatively charged amino acid (R/HGE/K) located in the first

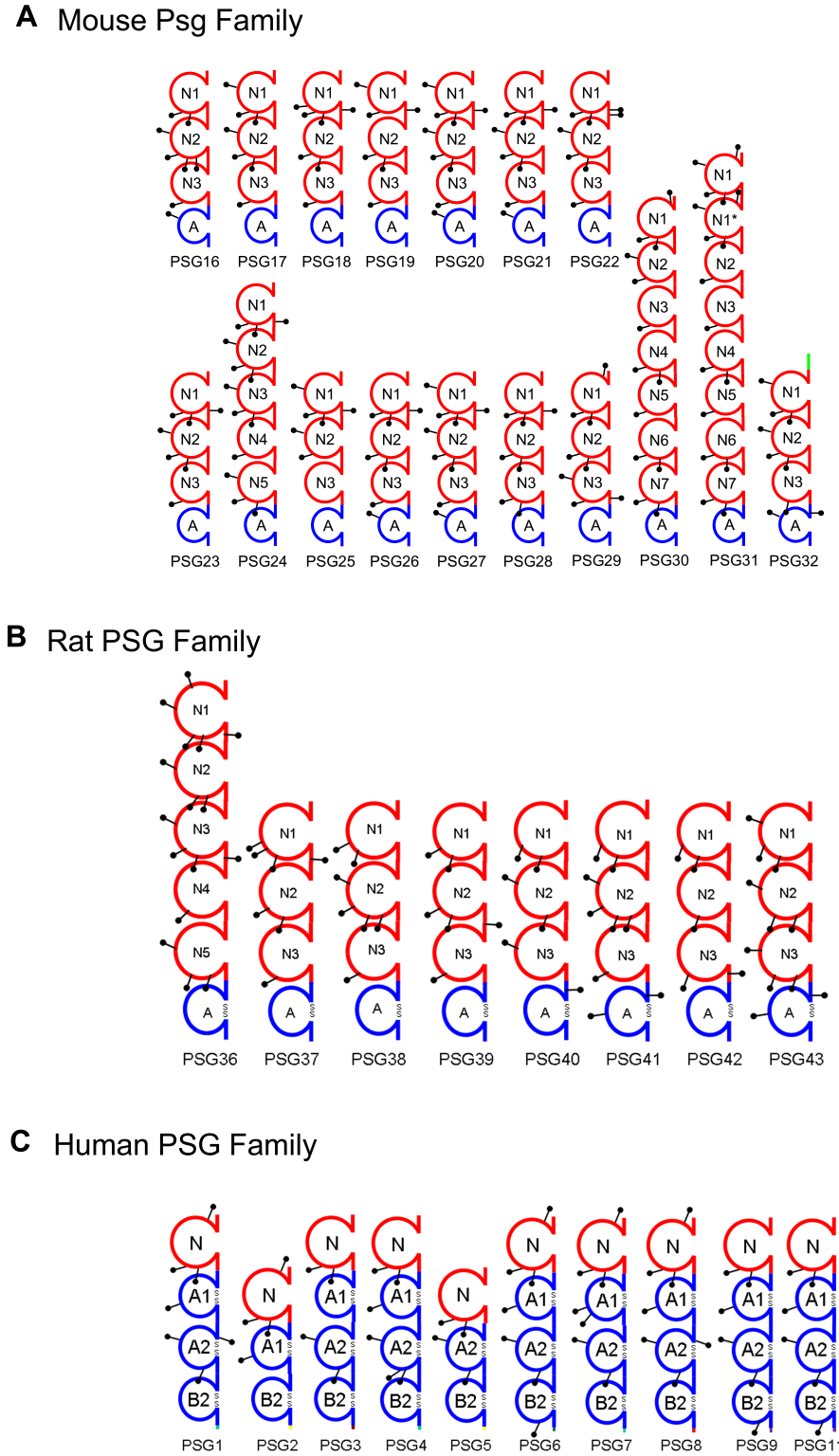


Figure 1.6: Schematic representation of Pregnancy-Specific Glycoprotein domain organisation in (A) Mouse *Psgs* are composed of 3 – 8 IgV-like N domains and one IgC-like A domain. The relative position of potential N-glycosylation sites indicated by lollipops [148]. (B) Rat *PSGs* are composed of 3 – 5 IgV-like N domains and one IgC-like A domain and (C) Human *PSGs* are composed of 1 IgV-like N domain, 2 or 3 Ig-C-like domains. (Not to scale) Human PSG5 also has a larger NA1A2B2 variant. (modified from http://www.carcinoembryonic-antigen.de/).

N domain [152, 106]. To asses the prevalence of conservation of this integrin binding motif across in PSGs across a variety of species, I aligned the N terminal IgV-like domain of Human, Chimpanzee, Baboon, Mouse and Rat PSG families. There is a high level of conservation of this motif throughout Old World primates and rodents species (Fig:1.7.). As predicted chimpanzees (*Pan troglodytes*) shared the highest level of RGD motif conservation within the 10 PSG protein coding genes that they posses (PSG1-11). The Baboon (*Papio hamadryas*) showed slight deviance from this motif, only one third of the 15 member PSG family (PSG56-PSG70) harbouring an RGD motif, whereas the remaining members have predominantly (QGD/RCD/RCH) motifs, with PSG70 possessing a unique PAE motif in the Baboon PSG family. This motif conservation is maintained but with a higher level of deviance in the rat (*Rattus norvegicus*), with the eight rat *PSG* (*PSG36-PSG43*) family members having (RH/GRA/EKD) motifs. Conserved RGD and RGD-like tri-peptide motifs in the majority of PSGs suggest that they may function like snake venom disintegrins, which bind integrins and disrupt tissue architecture of prey [153]. This integrin-binding motif that is thought to mediate interactions with the extracellular matrix [154] and immune cells [155]. This partial conservation of an evolutionary important integrin binding motif composed of RGD and RGD-like tri-peptides in primate and rodent N and N1 domains, therefore supports a role for these conserved motifs in PSG function [146].

To ensure that I was working with the correct sequences for these genes and transcripts, a current and detailed list of *PSG* accession numbers from the main publicly accessible genome browsers was compiled using all of the known *PSG* sequences and using the online NCBI BLAST sequence alignment tool. Each *PSG* accession number was checked and its corresponding sequence was BLASTed against entries from the three most commonly used sequence databases. A table of *PSGs* that correctly aligned to their corresponding sequences and annotations was constructed (Table 1.3.). This is an updated version of the accession table found in McLellan et al, 2005 [148]. Even though this current table is as up-to-date as possible, there are a few sequences that are not yet properly annotated, especially in the rat genome. The rat genome still requires completion as it has many regions of the genome yet

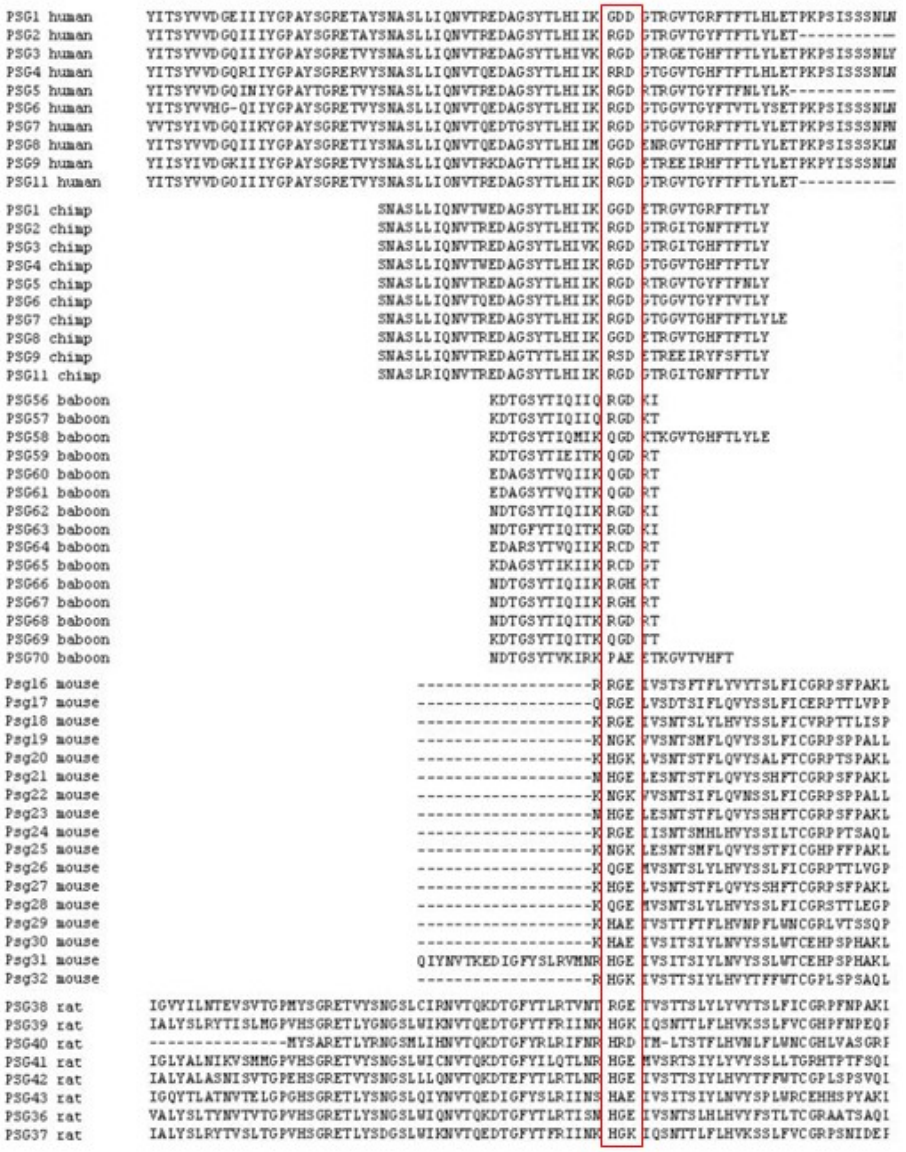


Figure 1.7: Functional conservation of integrin-interacting 'RGD'-like tri-peptide motif between species. ClustalW alignment of 'RGD'-like tri-peptide motif in N-terminal IgV-like domain of Human, Chimpanzee, Baboon, Mouse and Rat PSG families.

Table 1.3: Mouse, Rat and Human PSG accession numbers

PSG	NCBI mRNA	NCBI Protein	Coordinates	Ensembl mRNA	Ensembl Protein	CCDS
Psg16	NM007676.4	NP031702.3	17,074,040..17,098,971	ENSMUST00000071399	ENSMUSP00000071348	CCDS20861
Psg17	NM007677.2	NP031703.1	18,813,937..18,821,591	ENSMUST00000004655	ENSMUSP00000004655	CCDS20878
Psg18	NM011963.2	NP036093.2	18,345,808..18,355,007	ENSMUST00000003597	ENSMUSP00000003597	CCDS20869
Psg19	NM011964.2	NP036094.2	18,789,125..18,798,510	ENSMUST00000004657	ENSMUSP00000004657	CCDS20877
Psg20	NM054058.1	NP473399.1	18,674,366..18,685,992	ENSMUST00000108482	ENSMUSP00000104122	
Psg21	NM027403.4	NP081679.2	18,646,654..18,656,725	ENSMUST00000094793	ENSMUSP00000092387	CCDS20875
Psg22	NM001004152.2	NP001004152.1	18,718,090..18,727,248	ENSMUST00000051973	ENSMUSP00000050633	CCDS20876
Psg23	NM020261.4	NP064657.2	18,606,343..18,616,501	ENSMUST00000057810	ENSMUSP00000056586	CCDS20874
Psg24	NM054059.1	NP473400.1	17,150,400..17,163,231	ENSMUST00000108491	ENSMUSP00000104131	
Psg25	NM054060.1	NP473401.1	18,519,702..18,532,227	ENSMUST00000094795	ENSMUSP00000092389	CCDS20872
Psg26	NM001029893.1	NP001025064.1	18,474,582..18,484,149	ENSMUST00000094798	ENSMUSP00000092392	CCDS20871
Psg27	NM001037168.1	NP001032245.1	18,556,514..18,567,305	ENSMUST00000094794	ENSMUSP00000092388	CCDS20873
Psg28	NM054063.4	NP473404.3	18,422,536..18,432,055	ENSMUST00000019291	ENSMUSP00000019291	CCDS20870
Psg29	NM054064.3	NP473405.1	17,203,477..17,215,756	ENSMUST00000075934	ENSMUSP00000075320	CCDS20862
Psg30	NM028480.2	NP082756.1	17,713,252..17,761,121	ENSMUST00000081907	ENSMUSP00000080582	CCDS20863
Psg31	XM003084734.1	XP003084734.1	17,871,767..17,919,024	ENSMUST00000108490	ENSMUSP00000104130	
Psg32	NR002857.1		17,672,313..17,682,060	ENSMUST00000165490	ENSMUSP00000130117	
PSG36	NM012702	NP036834.2	77529236..77541770	ENSRNOT00000023663	ENSRNOP00000023663	
PSG37	NM019126	NP061999.1	78124961..78135217	ENSRNOT00000058000	ENSRNOP00000054809	
PSG38	NM012525	NP036657	77441013..77453826	ENSRNOT00000037086	ENSRNOP00000033854	
PSG39	XM218398	XP218398				
PSG40	NM021677	NP067709.1	77588947..77597830	ENSRNOT00000037132	ENSRNOP00000034066	
PSG41	NM001025679.1	NP001020850	77471004..77487104	ENSRNOT00000058060	ENSRNOP00000054868	
PSG42	NW047566.1		1108370..1118331			
PSG43	NW047566.1		1149404..1185428	ENSRNOT00000024155		
PSG1	NM001184825.1	NP001171754.1	43370613..43383871	ENST00000436291	ENSP00000413041	CCDS54275
PSG2	NM031246.3	NP112536	43568362..43586893	ENST00000406487	ENSP00000385706	CCDS12616
PSG3	NM021016	NP066296.2	43225794..43244668	ENST00000327495	ENSP00000332215	CCDS12611
PSG4	NM002780.3	NP002771	43696854..43709790	ENST00000405312	ENSP00000384770	CCDS46093
PSG5	NM002781.3	NP002772.3	43671895..43690688	ENST00000366175	ENSP00000382334	CCDS12617
PSG6	NM001031850.2	NP001027020.1	43406240..43422043	ENST00000187910	ENSP00000187910	CCDS33038
PSG7	NM002783.2	NP002774	43428284..43441330	ENST00000406070	ENSP00000421986	
PSG8	NM001130167.1	NP001123639.1	43256839..43269831	ENST00000404209	ENSP00000385869	CCDS46091
PSG9	NM002784	NP002775	43757435..43773682	ENST00000270077	ENSP00000270077	CCDS12618
PSG10	NR026824.1		43341149..43359870	ENST00000501199		
PSG11	NM002785	NP002776	43511809..43530631	ENST00000320078	ENSP00000319140	CCDS12614

unsequenced. This comprehensive accession table will aid in further studies of the PSG multi-gene families.

1.2.3 PSG function

The exact physiological functions of PSGs are not known. A conserved function between human and mouse *PSGs* has been proposed, due to conservation of structure and expression patterns. The fact that PSGs are synthesized by CTB and TGC, and are secreted from the outermost layer of the placenta that aggressively invades the uterine wall during placentation, implicates the PSG families role in structural modulation at the feto-maternal interface. Given that the PSGs are heavily glycosylated and the protein sequences are evolving rapidly, it is possible that PSGs function in a similar way to the ruminant PAGs. Indeed, the glycosylated PAG [156, 157] and PSG [148, 49] proteins are both implicated in immunological roles. The discovery that PSGs induce cytokines in human and murine macrophages has led to the consideration that human PSGs may function to modify maternal immune responses. Over a decade of published work has shown PSGs to be pro-angiogenic and immunomodulatory hormones that can directly induce various cytokines from several cell types in a cross-species reactive manner and suggests that PSGs exert an influence on cytokine polarization in pregnancy [158, 159, 160, 161, 162, 12, 163, 164, 13, 165, 166]. Table 1.4 outlines the published cytokine responses reported for individual PSGs and the responsive cell types. These anti-inflammatory cytokines promote a tolerogenic decidual microenvironment, and expression of the anti- inflammatory cytokines IL-10 and TGFβ1, by peripheral blood mononuclear cells (PBMC) and placenta has been associated with successful human pregnancy [167, 168, 169, 160].

Motrán *et al*, performed complex *in vivo* studies demonstrating that PSG alternatively activates antigen presenting cells which then polarize maternal T-cell differentiation to the 'less-damaging' Th2-type phenotype compatible with successful pregnancy [12]. Recently, the same group treated DC with PSG1, which promoted the enrichment of Th2-type cytokines, IL-17-producing cells, and Treg cells from CD4+ T cells from DO11.10 transgenic mice [166]. In parallel to their immunomodulatory

role, it has been indicated that some members of the murine and human *PSG* family may be involved in placental angiogenesis. It was found that PSG1 induces the formation of tubes by endothelial cells and members of the human and murine PSG induce the secretion of TGFβ1 and VEGF-A [170, 13, 171]. A possible role of PSGs in uteroplacental angiogenesis is further supported by the finding that incubation of endothelial cells with Psg22 resulted in the formation of tubes in the presence and absence of VEGFA [165]. PSGs role in vasculogenesis and angiogenesis may be required for the establishment and maintenance of the fetoplacental blood supply. It has also been shown that *PSG* genes can be categorized as early-responsive genes in cellular senescence models [172, 173], as all *PSGs* were upregulated in HeLa cells upon the addition of 5-bromodeoxyuridine in replicative senesence.

Recent work from our laboratory has demonstrated that PSG1 has other functions apart from cytokine induction, immune modulation and angiogenic stimulation (Table 1.4.) [158, 159, 160, 161, 162, 12, 163, 164, 13, 165, 166]. Moore and colleagues have shown that PSG1 exhibits a novel anti-thrombotic function, facilitated through the binding of many PSG1 domains to the α-2b-β-3 platelet integrin, which inhibits the platelet-flibrinogen interaction. Moore hypothesised that *PSGs*, evolved to prevent thrombosis at the placental surface or in the maternal circulation during pregnancy. PSG secreted into maternal blood would have to be at an elevated concentrations to counteract maternal fibrinogen which circulates in high levels in the maternal blood (2 mg/ml). This maybe an alternative explanation for the high PSG expression levels during pregnancy that was previously thought to be due to the Maternal-Foetal-Conflict theory (MFC) [174]. To date *PSGs* have been implicated in a variety of functions, from immune modulation, to angiogenic and anti-thrombotic molecules. Further study needs to be performed to discern whether all *PSGs* have a common function, or if individual *PSGs* perform specific functions, in accordance to their spatio-temporal expression patterns. One of only a few PSG receptors identified to date is the integrin-associated cluster of differentiation 9 antigen (CD9) receptor. CD9 is a member of the tetraspanin family, which is an important membrane protein with four transmembrane domains and two extracellular domains. Tetraspanin family

Table 1.4: Published cytokine responses for PSGs

PSG	Cytokines	Responsive Cell Types
1	IL-6 [160, 166], IL-10 [160], TGFβ1 [160, 12, 163, 13, 166], VEGF-A [13]	monocytes/macrophages, dendritic, endothelial, trophoblastic
6	IL-6 [160], IL-10 [160], TGFβ1 [160]	monocytes/macrophages
11	IL-6 [160, 163], IL-10 [160], TGFβ1 [160]	monocytes/macrophages
17	IL-6 [163], IL-10 [163], TGFβ1 [163]	macrophages
18	IL-10 [159]	macrophages
19	TGFβ1 [164]	macrophages
22	TGFβ1 [165], VEGF-A [165]	dendritic, natural killer
23	TGFβ1 [170] VEGF-A [170]	monocytes/macrophages, dendritic, endothelial, trophoblastic

members have been implicated in a variety of cellular and physiological processes, such as cell aggregation and motility, signalling, and fusion [175]. Dveksler and colleagues show that Psg17 and Psg19 bind to CD9 [161]. They also found that the amino acids involved in CD9 binding reside in the region of highest divergence between the N1-domains of murine *Psgs* [176]. In macrophages CD9 was found to bind the N1 domain of both Psg17 and Psg19. The interaction of Psg17 and CD9 was found to be necessary for the induction of secretion of anti-inflammatory cytokines [164]. Psg17 has also been shown to prevent sperm–egg fusion by interrupting the binding of CD9 to a ligand on the sperm surface [162]. Unlike mouse Psg17 and Psg19, human PSG does not require CD9 to induce cytokine production from mouse macrophages [164].

It was also recently discovered that the murine *Psgs Psg22* and *Psg23* and human PSG1 do not bind to CD9, but instead bind to heparan and chondroitin sulfate proteoglycans (HSPG) [171, 165]. Specifically, PSG1 binds syndecans 1-4 and glypican-1 on the surface of cells [171]. Proteoglycans (PGs) consist of a protein core and covalently attached glycosaminoglycan (GAG) chains [177]. The syndecans are considered hybrid PGs since they contain mixtures of the two major types of GAG chains found in animal cells, heparan sulfate and chondroitin sulfate. There are four members of the syndecan family, syndecan-1 (CD138), syndecan- 2 (fibrogycan), syndecan-3 (N-syndecan), and syndecan-4 (ryudocan or amphiglycan). The other major family of membrane PGs comprises the glypicans (-1 to -6), which contain glycosylphosphatidylinositol anchors instead of a membrane-spanning segment [178].

This binding of PSG1 to GAGs on the surface of endothelial cells mediates tube formation implicates a PSG-GAG interaction that mediates certain PSG angiogenic functions [171]. The finding that syndecan-1 regulates two critical integrins in angiogenesis, α-v-β-3 and α-v-β-5 [179], further supports the role of syndecans in angiogenesis. The presence of multiple possible PSG receptors, suggests multiple functions for *PSGs* interacting through these different receptors. The identification of receptors for every *PSG* in this multigene family will help our understanding of individual *PSG* function, and about the function of the family as a whole.

1.2.4 PSG expression

As stated previously, human *PSGs* are tightly linked on the long arm of chromosome 19 and it has been shown that they are coordinately expressed in the placenta [180]. Individual *PSG* member expression study is demanding due to their high degree of sequence identity and the lack of specific antibodies for each PSG protein. Human *PSG* transcripts and proteins increase in trophoblast cells undergoing differentiation [135, 181] and are detectable until term. They are secreted by the syncytiotrophoblast and are detected 3-4 days after fertilization, concordant with blastocyst implantation [182]. PSG1 has been identified as the most active transcript up-regulated (70 fold) during the *in vitro* cell differentiation of CTB to syncytiotrophoblast [183]. Specific transcripts for *PSG1, PSG3, PSG5, PSG7* and *PSG9* genes were detected in differentiated JEG-3 and CTBs while they were undetectable or had low expression level in undifferentiated cells [181]. It has been reported that all human *PSG* mRNAs can be detected in placenta at different levels, although due to the similar nature of PSGs at the amino acid level, it is difficult to determine the protein expression pattern because of high cross-reactivity with monoclonal Abs [180, 184, 185]. Present data suggest that the whole *PSG* locus is activated in CTB that differentiates into the syncytium pathway, although they reach different abundance levels. It was initially hypothesised that *PSGs* were expressed exclusively in the placenta [129], but it has been described that human *PSGs* are also expressed in the non-pregnant state including breast cancer, choriocarcinomas, peripheral blood cells, and bone

marrow cells [186, 187, 188]. High expression of *PSGs* in breast cancer have been correlated with a poor prognosis [187]. It is interesting to note that syncytium-like trophoblast cells express very little *PSG9* mRNA and, conversely, up-regulation of *PSG9* expression, but not any other *PSG* member, was found in colorectal carcinogenesis [189, 190].

There are 17 mouse *Psg* family members with different expression levels at different stages of development, with Psg22 has been identified as an earliest *Psg* expressing gene in the mouse placenta [146, 49]. *Psg22* mRNA was detectable around the embryonic crypt on E5.5 and became most prominent at E10.5 coinciding with placental formation, indicating that TGC are the main source of *Psg22* during the early develoment of the foetal-maternal interface [165]. Employing qRT-PCR with primers that amplify all mouse *Psgs*, it was found that in TGC, there is increased *Psg* expression between E8 and E11 with *Psg* transcript levels doubling from E8 to E9 and from E9 to E10. In EPC, there is a fivefold increase in *Psg* transcript levels between E9 and E11. However, absolute levels in the EPC are low, with E10 TGC having approximately sixfold higher levels than E10 EPC [49]. Moore and colleagues also discovered that *Psg22* is the most abundant transcript in the first half of pregnancy, with *Psg16, Psg21* and *Psg23* accounting for 90% of transcript abundance in the second half of pregnancy [191]. The early expression of *Psg22*, together with its pro-angiogenic effects suggest that this protein may play an important role modulating the ability of DC and NK cells to induce the early vascular adaptations required for successful implantation and placentation [165]. *Psg* gene expression data in mouse pregnancy implies that different family members show different expression levels between E11 and E18, implying the possibility of divergent functions of individual *Psgs* in the mouse [148]. Kromer *et al*, [129], tested for *Psg17, Psg18* and *Psg19* mRNA and found that murine *Psg* transcripts are detectable by means of RT-PCR analysis in the placenta and the pooled tissues of embryo but not in adult tissues, including kidney, lung, testis, ovary, liver, brain, thymus, heart, and spleen. Although non-placental murine *Psg* expression was realised when *Psg18* was found to be highly expressed in the follicle-associated epithelium (FAE) overlaying Peyer's patches (PPs) [192] implicating *Psg18*

in the modulation of the mucosal immune system, and a *Psg16* brain specific transcript was also recently detected [193]. Further study needs to be performed regarding the expression of other murine *Psg* family members in non-placental cell types, including the gatrointestinal tract (GIT).

1.2.5 PSG regulation

Despite such extensive knowledge about the structure and function of the murine *Psg* genes, relatively little information is available about regulation of the murine *Psgs* at the transcriptional level or the promoter regions that infer this regulation. It was established that the biosynthesis of PSGs is mainly regulated at the transcriptional level, with an increase in their expression during placental development [194]. Very little information has been generated concerning the regulatory mechanisms that control the individual murine *Psgs*. A comprehensive review of the literature regarding the regulation of human and mouse *PSGs* can be viewed in Table 1.5. What information is known regarding *PSG* regulation is mostly concerned with Human *PSGs*, and more precisely *PSG5*, as the 5'-flanking sequence of the *PSG5* gene has been characterized and used as a model for studies of *PSGs* regulation due to the strong homology of promoter sequences among the different family members [195].

The Human *PSG* genes are extremely similar and that similarity extends to their

Table 1.5: Published regulators of *PSG* expression

Regulator	***PSG***	**Responsive Cell Lines**	**References**
Cpbp	*PSG5*	Jeg3	[196]
Klf4	*PSG5*	Jeg3, HeLa	[86, 197]
Klf6	*PSG3, PSG5*	Jeg3	[86, 198]
Sp1	*PSG5*	Jeg3, HP-A1	[189, 199, 197]
RxR	*PSG5*	Jeg3,	[200]
XBP1, IRE1a	*Psg18, Psg28*	SM-10, MEF	[201]

putative control regions [189]. Human *PSGs* do not have conventional promoters, as promoters of human *PSG* genes are highly homologous and lack any obvious TATA-box, typical Initiator elements, or large GC-rich sequences [202, 195]. Human *PSGs* have been defined to possess minimal promoter regions, spanning from -172 to -34 bp

[189]. Previous work on the Human *PSG* promoters has shown that *PSG5* expression is dependant on a functional ubiquitous specificity protein 1 (Sp1) binding site located in the minimal core promoter element of all human *PSGs* (-140 to -147). This SP1 site has been shown to activate *PSG5* promoter constructs, and is coexpressed with *PSG5* in human placental villi, particularly the syncytiotrophoblast layer, stressing its important role in the regulation of *PSG5* [199]. It has also been shown that in general, activation of the minimal basal promoter activity in *PSG5* in the HP-A1 cells requires minimal promoter lengths (172 bp upstream of the transcription start site) and the presence of Sp1 or Sp1-like elements and that the RARE motif is involved not only in basal promoter activity but also in *PSG* activation upon trophoblast differentiation.

Using gene promoter-reporter transfections and X-ChIP assays, Blanchon *et al*, demonstrated that Kruppel-like factor 4 (KLF4) is an activator of the *PSG5* promoter by binding to a KLF consensus like binding which includes the Core Promoter Element region (-147/-140) [197]. Furthermore, linking previous data showing the binding of Sp1 transcription factor to a GT-box (-443/-437) and co-transfection assays with KLF4 and Sp1, they were able to comprehensively demonstrate the robust combined activity of these two factors on the *PSG5* promoter. This transcriptional regulation of *PSG5* by KLF4 and the Sp1 transcription factor is synergistically co-activated by KLF4 and Sp1, and has been shown to require two intact DNA regions: the -148/-133 promoter sequence (TS1 site bearing the CPE-box) for KLF4 and the -443/-437 (GT-box) upstream element, for Sp1 [197]. The interaction of KLF4 with a house-keeping transcription factor such as Sp1 to regulate the placental-specific expression of *PSG5* is reminiscent of situations previously described in which tissue-specific and ubiquitous transcription factors interact to control specific gene expression [203]. Racca *et al*, [198], have also shown that Kruppel-like factor 6 (KLF6) is also involved in the activation and regulation of *PSG3* and *PSG5* promoters in Jeg3 cells, further supporting the role of the KLF family in *PSG* regulation. This group demonstrated increased expression of both human hCG and *PSG* genes using overexpression studies of KLF6.

Further investigation of this core promoter element (CPE), has revealed that this site partially overlaps a putative Retinoic acid Response Element (RARE) site,

conserved between positions -161/-145 in PSG genes, indeed, this RARE/CACCC-box composite element is almost identical in the 11 human *PSG* genes [200]. This implicated the existence of an RXRa-mediated pathway leading to PSG gene activation through this conserved RARE motif. Retinoic Acid (RA) is the active derivative of vitamin A (retinol), and exerts its effects through two families of receptors, retinoid acid receptors (RARs) and retinoid X receptors (RXRs), which act as ligand-inducible transcription factors [204]. Some observations suggest that RA may be involved in placental development. For example, RARs and RXRs show localized expression in the placenta [32], and RA promotes TSC toward the TGC fate [101]. Lopez-Diaz *et al*, [200], have demonstrated that RXRα does bind to *PSG5* CPE and that 9-cis Retinoic acid induces *PSG5* expression in JEG-3 cells, thus, it seems possible that cells committed to differentiate into syncytiotrophoblast are able to respond efficiently to RXR signaling, leading to increased transcriptional response of PSG genes [200]. The region (-178/-49) of the *PSG3* promoter contains several consensus DNA binding sites, among them are a RARE motive and a putative binding site for the Ets-family transcription factor GABP. It was shown using luciferase assays that the RARE binding site is required for basal promoter activity while the GABP binding site is involved in the induction of *PSG3* transcription during differentiation [181].

Expression of *PSG* genes is regulated by the interaction of transcription factors with positive and negative DNA elements in the *PSG* promoters as shown by [189]. Transcriptional control was further investigated in primary CTB cultures indicated the presence of a functional repressor element located upstream -251 nt, as it had been described for *PSG5* in non-placental cells [181, 205]. *PSG5* regulation is not only mediated by transcriptional level control via DNA binding factors, [205], Panzetta-Dutari *et al*, have described *cis* and *trans* acting negative elements, that function in repressing *PSG5* transcription, irrespective of the cell type. All *PSG* family members were found to be clearly up-regulated by addition of 5-bromodeoxyuridine in HeLa cells. Likewise, all *PSG* family members were clearly up-regulated in normal human fibroblasts during replicative senescence. Promoter analysis of the *PSG1*, *PSG4*, and *PSG11* genes in HeLa cells did not possess a *cis*-regulatory element

responsive to 5-bromodeoxyuridine in their 50bp-flanking sequences. These results suggest that the *PSG* genes are regulated at a level of higher order chromatin structure [173].

Inositol requiring enzyme-1a (IRE1a) is an endoplasmic reticulum (ER) located transmembrane RNase whose activation leads to the production of the transcription factor X-box binding protein 1 (XBP1) [206]. Oikawa *et al*, have recently shown that following treatment with thapsigargin, a typical ER stressor that activates the IRE1a–XBP1 pathway, or using overexpression of wild type IRE1a or XBP1, both *Psg18* and *Psg28* were upregulated in SM-10 cells [201]. Through *Psg28* promoter region deletion constructs, they identified two important regions whose individual deletion reduced promoter activity, firstly, the (-500/-480) upstream region, and the second was positioned in the (-180/-140) upstream region. As these two regions did not contain any previously identified XBP1-responsive elements, nor XBP1 binding sites, it is likely that XBP1 up-regulates the *Psg28* promoter in an indirect manner, possibly through up-regulation or activation of intermediate factors [201]. This is one of the first molecules to be implicated in the regulation of murine *Psgs*.

This section has discussed the multitude of genes and mechanisms involved in the regulation of *PSG* expression, the majority of which, do so at the transcriptional level. To date, there has been little data generated in the literature published concerning the epigenetic control of this multigene family, although it has been suggested that a fine-tuning complex mechanism that may include specific long-range acting chromatin factors, transcriptional regulation and transcript stability controls the expression of each *PSG* gene member [181]. In the next section I will discuss the ways in which epigenetic regulation modulates gene transcription, especially through the use of non-coding ribonucleic acid (ncRNA) transcripts.

1.3 Chromatin

1.3.1 Epigenetics and the role of chromatin

Epigenetics has a been the focus of research in recent years, as it is concerned with the contextual information that is superimposed on the relatively stable underlying genomic sequence, by the modification of DNA (and ribonucleic acid (RNA)) and the modulation of chromatin structure [207]. Gene regulation through epigenetics is essential for producing variance of cell types during mammalian development, and is crucial for sustaining the stability and integrity of the expression profiles of different cell types [208]. Chromatin is the state in which DNA is packaged within the cell through the association with histone proteins [209]. The inheritance of chromatin states such as "active" (euchromatic) or "silent" (heterochromatic) domains forms the foundation of epigenetics [210]. The nucleosome is the fundamental unit of chromatin and it is composed of an octamer of the four core histones (H3, H4, H2A, H2B) which are wrapped around 147 base pairs of DNA. The core histones are primarily globular except for their N-terminal "tails," which are unstructured. The amino acid sequence of these N-terminal tails is highly evolutionary conserved. This level of conservation implicates a selective force which maintains the sequence of the N-termini. This conservation is due to the tails undergoing multiple post-translational modifications. These modifications subsequently can modulate chromatin structure [211]. Chromatin architecture is altered by methylation of the DNA and by various types of modifications to histones (the so-called 'histone code'), including compound patterns of acetylation, methylation, phosphorylation, ubiquitinylation, sumoylation, ADP-ribosylation, carbonylation, de-imination and proline isomerization at various residues, (reviewed extensively by [212, 208, 213]). The remarkable intricacy of covalent histone modifications is exacerbated by the presence of histone variants in numerous organisms. These histone modifications convey additional possibilities for the cell to diversify the overall composition of the nucleosome and its covalent modification potential [214]. In eukaryotic organisms, chromatin is involved in many different processes, from development, cognition, ageing to disease progression.

Understanding how chromatin directs gene expression remains to be an important focus of research [215].

1.3.2 Long Noncoding RNAs (lncRNA)

Advances in technology have assisted in the high resolution analysis of the human and mouse transcriptomes [216], illustrating that the transcriptome of the mammalian genome is much larger than originally thought [217, 218]. Proteins and related protein-coding genes have been at the centre of biological research for years. Nonetheless, the development of bioinformatical methods and advanced RNA sequencing technology for compiling the transcriptome, has illustrated that besides protein-coding genes, the majority of the mammalian genome is transcribed, and many noncoding RNA (ncRNA) transcripts contribute to a variety of biological roles [219, 220]. The discovery of extensive transcription of large RNA transcripts that do not code for proteins, termed long noncoding RNAs (lncRNAs), provides a new insight in the pivotal role of RNA in gene regulation [221]. Advancing technologies, including RNA-Seq, are not confined to the identification of protein-coding RNA transcripts, and have facilitated in the discovery of many novel lncRNA transcripts. These transcripts were generally believed to be "junk," but current studies suggests that the majority of these RNAs are essential in regulating gene expression at various levels [222]. Currently, these lncRNAs are defined as RNA genes, which are larger than 200 bp but do not appear to have coding potential. However, the size cutoff clearly distinguishes lncRNAs from small regulatory RNAs, including micro RNAs (miRNAs), transfer RNAs (tRNAs), or piwi RNAs (piRNAs). LncRNAs have also been classified using the anatomical characteristics of their gene loci. For instance, lncRNAs are often defined by their position relative to neighbouring protein-coding genes. (Fig:1.8. A-D) shows the four main lncRNAs that have been described to date. Antisense lncRNAs (A), are lncRNAs whose transcription commences inside or 3' of a protein-coding gene, are transcribed in the antisense direction of protein-coding genes, and share an overlap of at least one coding exon. Intronic lncRNAs (B), are lncRNAs whose transcription commences inside of an intron of a protein-coding gene in either

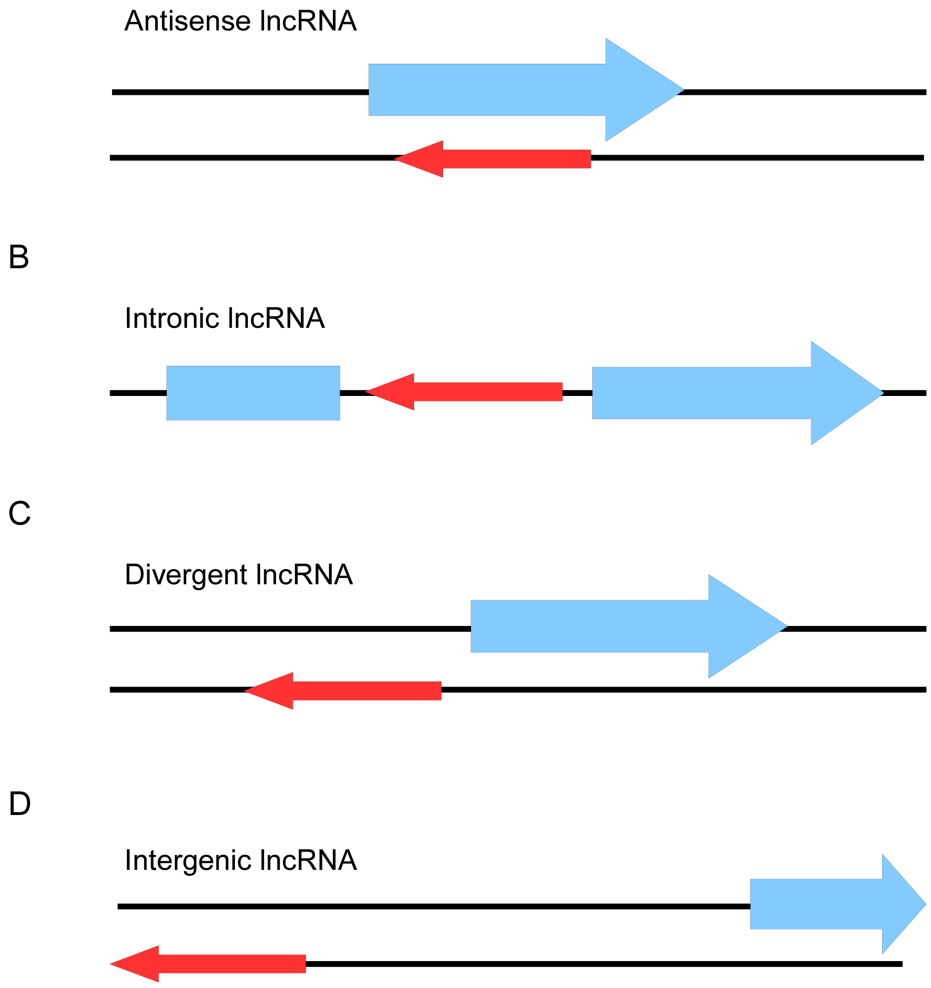


Figure 1.8: Anatomy of long non-coding RNA (lncRNA) loci. (A) Antisense lncRNA - lncRNA sequence overlaps with the antisense strand of a protein coding gene. (B) Intronic lncRNA - lncRNA sequence is derived entirely from within an intron of another transcript. (C) Divergent lncRNA - lncRNA sequence is located on the opposite strand from a protein coding gene whose transcription is initiated less than 1000 base pairs away. (D) Intergenic lncRNA - lncRNA sequence is not located near any other protein coding loci. Modified from [221, 223].

direction and terminate without overlapping exons. Bidirectional lncRNAs (C), are transcripts that initiate in a divergent or bidirectional fashion from the promoter of a protein-coding gene; and although not exactly defined, generally initiate transcription within a few hundred nucleotides of the neighbouring promoter. Intergenic lncRNAs (D) which are sometimes termed large intervening noncoding RNAs or lincRNAs, are lncRNAs that possess separate transcriptional units and are over 5 kb from their protein-coding gene neighbours [221]. At present, lncRNAs are defined by their

size and anatomical properties, as previously stated but are also characterised by their protein coding potential. Whether an RNA transcript functions by coding for protein in any of its 3 frames is fundamental to the definition of lncRNA [221]. A new method of characterising lncRNAs, termed guilt by association, which associates protein-coding genes and lnRNAs that possess concordant expression patterns and are therefore presumed to be co-regulated, has enabled a comprehensive understanding of lncRNAs [224]. Employing gene-expression analyses, this approach identifies protein-coding genes and regulatory pathways that correspond with the lncRNA under investigation. Using data generated from these concordantly expressed protein coding genes, the functions and regulatory mechanisms of the lncRNA is inferred. Expression patterns of lncRNAs are associated with numerous key cellular processes, including immune responses [225], pluripotency [226], and regulation of the cell cycle [227]. To date, it has been shown that almost a third of lincRNAs associate with chromatin-modifying complexes [228]. This association of lncRNA with ribonucleic-protein complexes is the mechanism in which they exert their influence on the regulation of gene expression [221]. A recent study has revealed a number of interesting properties of lncRNAs. These properties include being predominantly positioned neighbouring developmental regulators, enhancement of tissue-specific expression patterns, possessing many orthologous Large intergenic non-coding RNAs (lincRNAs) between human and mouse, and the abundant presence of lincRNAs in in genetic loci that are associated with genetic traits but contain no protein-coding genes [229].

Despite the extensive data being generated concerning expression of these lncRNAs, the functional roles for lncRNAs have remained mostly elusive. lncRNAs were once thought of as the "dark matter" of the genome, due to our lack of functional knowledge regarding these RNA transcripts [219]. Recently, a number of examples have arisen to suggest that the co-transcription of non-coding transcripts influences neighbouring gene transcription. These lncRNAs have been shown to be involved in both repression and enhancement of gene transcription through many different mechanisms. To date, the known functions of lncRNAs have been reviewed

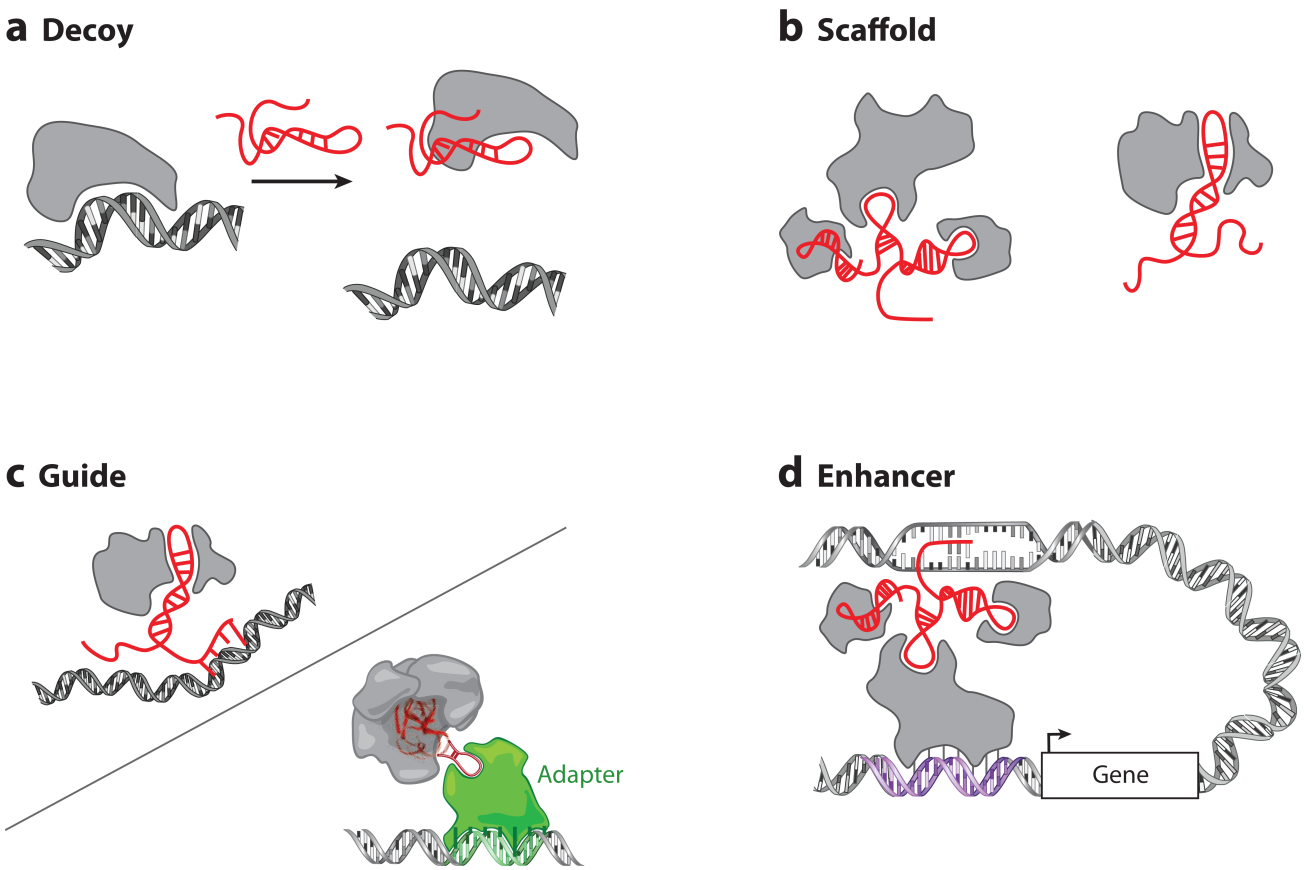
extensively [219, 221]. They have reviewed four mechanisms in which lncRNAs regulate gene transcription. The four proposed mechanisms in which lncRNAs are hypothesised to work are shown (Fig:1.9. A-D).

LncRNAs can act as decoys that act as a sink to remove DNA-binding proteins, such as transcription factors, thus repressing neighbouring gene expression. One example of lncRNAs acting as decoys is the recent example, PANDA, which is induced in a p53-dependent manner. PANDA interacts with the transcription factor NF-YA to limit expression of pro-apoptotic genes implicating lncRNAs in the control of cell growth [227]. Other examples of decoy lncRNAs include TERRA which regulates and protects chromosome ends [230], and MALAT1 which is involved in alternative splicing regulation mediated through splicing factor phosphorylation [231]. LncRNAs can act as scaffolds, to facilitate in the formation of protein complexes or to bring these proteins in proximity to the loci. HOTAIR is an example of a scaffold lncRNA which regulates epigenetic states, through binding of both PRC2 and LSD1-CoREST complexes via specific RNA domains [224]. TERC is another scaffold lncRNA which regulates telomerase catalytic activity through the formation of protein complexes that are essential for telomerase function [232]. LncRNAs can act as guides to recruit proteins such as chromatin modification enzymes to the loci, through specific RNA-DNA or RNA-protein interactions. These guide lncRNAs, such as Xist, Air and Kcnq1ot1 [233, 234] are involved in dosage compensation and imprinting. Another example of guide lncRNAs is lincRNA-p21, which acts as a repressor of transcription, mediated through interactions with hnRNP-K which results in p53-dependent transcriptional responses to DNA damage [225].

lncRNAs are involved in enhancer-regulating gene activation (eRNAs), through chromosome looping in which cases they may interact directly with distal genomic regions. One activity-regulated neuronal enhancer was independently identified as an enhancer that drives the activity-regulated transcription of *arc/arg3.1*, a gene that regulates synaptic function [235, 236, 237]. This *arc* enhancer which is located 7 kb upstream of the transcriptional start site (TSS), is necessary to drive activity-regulated *arc* transcription [238, 239]. HOTTIP is yet another example of

lncRNAs that are associated with gene activation and euchromatin. HOTTIP is located on the distal 5′ end of the HOXA gene cluster and binds with the WDR5 protein to activate histone H3 lysine 4 trimethylation. Through chromosomal looping of HOTTIP, thus bringing HOTTIP in the proximity of a number of HOXA cluster genes. This maintains histone H3 lysine 4 trimethylation and facilitates target gene activation [240]. This implicates eRNAs in the regulation of genes that are responsible for a number of essential developmental processes. Further research is needed to gain a comprehensive understanding of the roles and mechanisms of these complex lncRNAs *in vivo*.

Figure 1.9: Models of long non-coding RNA (lncRNA) mechanisms of action. (a) lncRNAs can act as decoys that act as a sink to remove DNA-binding proteins, such as transcription factors. (b) lncRNAs can act as scaffolds, to facilitate in the formation of protein complexes or to bring these proteins into proximity of the loci. (c) lncRNAs can act as guides to recruit proteins such as chromatin modification enzymes to the loci, through specific RNA-DNA or RNA-protein interactions. (d) lncRNAs are also involved in enhancer-regulating gene activation (eRNAs), through chromosome looping in which cases they may interact directly with distal genomic regions. Modified from [221].



1.4 Summary and Aims

Since the first discovery of the PSG family in the serum of normal pregnant women in 1970, there has been extensive research carried out in the expression, regulation and function of these complex multigene families, in a variety species possessing hemochorial placentation. Nevertheless, the molecular mechanisms implicated in their specificity of placental expression and their trophoblastic regulation are still poorly understood. An extensive analysis of murine *Psg* expression patterns in trophoblast lineages, particularly *Psg22*, has not been carried out. Furthermore, it is still unknown whether the murine *Psgs*, which exhibit differing RGD-like integrin binding motifs, possess the same functions and regulation mechanisms. Therefore, the aims of this thesis are:

1. Define and comprehensively map the rodent *PSG* loci
2. Understand expression of murine *Psgs* in TGCs, and trophoblastic lineage tissues, especially *Psg22*, which has been shown to have the highest expression levels of *Psgs* in the first half of pregnancy
3. Investigate the functions of Psg22 protein *in vitro*
4. Determine the regulatory mechanisms involved in the expression of *Psg22*

Chapter 2

Materials and Methods

2.1 Materials

All chemicals used were purchased from Sigma Aldrich unless otherwise stated. All restriction enzymes were purchased from New England Biolabs (NEB, Ireland). T4 DNA ligases were purchased from New England Biolabs. High Fidelity Phusion 2 Hot Start Thermostable DNA Polymerase was purchased from ThermoScientific. Plasmid DNA isolation, gel purification and nucleotide clean-up kits were purchased from QIAGEN. All oligonucleotide primers for polymerase chain reaction (PCR) were purchased from MWG Eurofins (Eurofins MWG Operon, Germany). All bacterial media constituents were purchased from Sigma Aldrich (Sigma-Aldrich, Ireland). Perfectly Blunt Cloning kit and Escherichia coli bacterial strains used were both purchased from Novagen (Merck, Germany). All plastics and mammalian tissue culture materials were purchased from Starstedt. DNA ladders and protein markers were purchased from New England Biolabs.

2.2 Bioinformatics

All PSG sequences (genomic, coding sequences (CDS), and amino acid (aa)) were taken from publically accessible genome browsers; National Centre of Biotechnology

Institute (NCBI) (http://www.ncbi.nlm.nih.gov/), University of California, Santa Cruz (UCSC) (http://genome.ucsc.edu/), and the Ensembl Genome browser (http://www.ensembl.org/index.html). Using these databases, a comprehensive accession table of all known rodent and human PSGs was compiled. I compiled available sequence data for all rodent PSGs and used sequence alignment software tools to locate the entire mouse and rat *PSG* gene families on their respective loci. Sequence alignments were performed using the online NCBI BLAST sequence alignment tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and the online ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html) alignment software. Also using ClustalW alignment software and the MEGA Molecular Evolutionary Genetics Analysis software MEGA5 (http://www.megasoftware.net/), I aligned individual species *PSG* families for mouse, rat and human *PSG* coding sequences (CDS) and constructed Phylogenetic trees (Neighbour-joined pairwise comparison phylogenetic trees). The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model [241]. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed [242]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [242]. All major branches yielded values of 95–100%. The scale bar represents 0.1 nucleotide substitutions per site.

Open Reading Frame (ORF) predictions were performed using the online ORF prediction softeware. This ORF prediction softeware is available from NCBI, (http://www.ncbi.nlm.nih.gov/projects/gorf/gorf.html). *PSG* promoters were analysed for putative transcription factor binding sites along a 2 kb region upstream of the Transcriptional Start Site (TSS) using the online MatInspector programme (Genomatix Software Suite, Germany) (http://www.genomatix.de/). Protein domain structure prediction was carried out using the online SMART (a Simple Modular Architecture Research Tool) software (http://smart.embl-heidelberg.de/). All primers were designed using the online Primer-Blast software (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) unless otherwise stated. Primer

analysis and the potential of secondary structures in primers designed was assessed using the online NetPrimer software (http://www.premierbiosoft.com/netprimer/). shRNA oligonucleotides were designed using the PSICOOLIGIOMAKER1.5 software programme which is available from the Jacks Lab (http://web.mit.edu/jacks-lab/protocols/pSico.html) to design the target sequences.

2.3 Molecular Biology

DNA purifications, agarose gel electrophoresis, cloning, PCR, RT-PCR, qRT-PCR and bacterial transformation were performed using standard molecular biology techniques or according to the relevant kit instructions.

2.3.1 Mice and Tissues

Mouse tissues were obtained from the Biological Services Unit, University College Cork. Mouse strains used were CD1, C57BL/6J, 129/Sv. Embryonic (E) stage refers to the gestational age of the embryo. The morning on which the vaginal plug was found is counted as day one (E1) of gestation. Human term placenta and human esophageal RNA (Ambion® FirstChoice® Human Total RNA Survey Panel, Life Technologies, AM6000) was kindly provided by Aine Fanning, Dept. of Medicine, UCC.

2.3.2 Cell culture

E2 mouse embryos were flushed from mated superovulated CD1 uterine horns as described elsewhere [243]. Embryos were placed in M2 medium microdots under mineral oil in embryo culture dishes. After two days M2 medium was replaced with M16 medium. Embryos were maintained in M16 medium under mineral oil until they reached E5 blastocyst stage. Some embryos were harvested at E5 and the rest were cultured until E11 to allow for blastocyst outgrowths to form, then harvested for RNA extraction and cDNA synthesis. The RAW-246.7 murine macrophage cell line was maintained in T75 flasks in Dulbecco's modified Eagle's medium (DMEM)

containing 10% fetal bovine serum; 1% penicillin/streptomycin and 1% L-glutamine. The human THP-1 monocytic cell line was maintained in T75 flasks in RPMI-1640, 0.05 mM 2-mercaptoethanol; 10% fetal bovine serum; 1% penicillin/streptomycin and 2 mM L-glutamine. JAR human choriocarncinoma cells were maintained in T75 flasks in DMEM containing 10% fetal bovine serum; 1% penicillin/streptomycin and 1% L-glutamine. Rat choriocarcinoma Rcho-1 trophoblast cells were a kind gift from M.J. Soares, Kansas. They were maintained in a subconfluent condition in T75 flasks with an RPMI-1640 medium supplemented with 10% fetal bovine serum, 50 uM 2-mercaptoethanol, 1 mM sodium pyruvate, 2 mM glutamine, 100 U/ml of penicillin, and 100 µg/ml of streptomycin. Differentiation was induced by growing the cells to confluence and subsequently replacing the 10% FBS supplementation with 1% donor horse serum. 3T3 mouse embryonic fibroblast cells were maintained in T75 flasks in DMEM containing 10% fetal bovine serum; 1% penicillin/streptomycin and 1% L-glutamine.

Primary mouse embryonic fibroblasts (MEF) were derived from E12.5 C57BL/6J mouse embryos as described in [244], embryos were dissociated and then trypsinized to produce single-cell suspensions in T175 flasks. These single cell suspensions were expanded, leaving only primary MEFs remaining. These were aliquoted and frozen at -80°C until ready to be used to harvest conditioned MEF medium, to be used as a trophoblast stem cell medium supplement. MEFs were maintained in DMEM containing 10% fetal bovine serum, 1% penicillin/streptomycin and 1% L-glutamine. Primary MEF cells were grown to confluence in T75 cell culture flasks and then treated with 10 µg/ml mitomycin C (MMC) for 3 hours. MMC containing medium was removed, cells were washed three times in PBS, and then fresh complete DMEM medium was added to the cells. MEF cells were used to condition medium for three days and then the medium was harvested, sterile filtered with a 0.2 µm syringe filter and then stored at -80°C until ready to use.

In the mouse, TSC are readily obtained by culturing cells from the extraembryonic ectoderm of implanting embryos or from outgrowths of cultured blastocysts and can be maintained in a pluripotent state in culture in the presence of

FGF4, heparin, and fibroblast conditioned medium, without which these cells begin to differentiate into the various trophoblast subtypes [21, 245]. Trophoblast stem cell lines (TS-GFP and TS-R26) were kindly donated to us by Dr. Myriam Hemberger (Babraham Institute, Cambridge, UK). The TSC line (TS-EXE) was kindly donated by Dr. Tilo Kunath (University of Edinburgh). TSC were maintained as described previously [246, 21]. TSC cell medium contained RPMI 1640 medium supplemented with 20% fetal bovine serum; 2 mM L-glutamine; 1 mM sodium pyruvate; 100 mM 2-mercaptoethanol; 50 U/ml penicillin and 50 µg/ml streptomycin. TSC seeded in T75 flasks were kept in an undifferentiated state using 70% fibroblast conditioned medium (FCM) +FGF4H medium. This medium constituted of TSC medium (described above) containing 70% MEF-conditioned medium and 25 ng/ml FGF4 and 1 µg/ml heparin. TGC were differentiated from TSC by culturing undifferentiated TSC in TSC medium without 70FCM+FGF4H for 6 days [21]. The transcriptional induction of *Pl2*, a prolactin family members that is only expressed in TGC, confirmed differentiation toward the TGC lineage [247].

The Freestyle™ 293-F cells were grown in suspension in Freestyle™ 293 Expression Medium, by shaker culture, in the presence of Antibiotic Antimycotic Solution (AAS) at 10 ml/L for two passages, and then Freestyle™ 293-F cells were grown without AAS to a density of 1 x 10^6 cells/ml. Freestyle™ 293-F cells were maintained at 37°C in a humidified 8% CO_2 shaking incubator. All reagents were obtained from Sigma-Aldrich, UK, unless otherwise stated. Cells were maintained in a humidified 5% CO_2 incubator unless otherwise stated, and were split regularly to ensure exponential rates of growth.

2.3.3 Cell Transfections and Treatments

All cell transfections were carried out using Lipofectamine 2000 (Invitrogen, 11668-019), and OptiMEM (Invtirogen, 31985-062) as per manufacturers instructions. Cells were grown to 90% confluency in 24 well plates and were cultured for 48 hours post-transfection. *Psg22* shRNA vectors were transfected into subconfluent TSC, using Lipofectamine 2000, in serum free medium for 6 hours, after which the medium

was changed to the differentiating GC medium and these transfected TSC cells were differentiated into TGC as described above. Untransfected TSC, an empty pSicoR vector, and a nonsense shRNA pSicoR construct were used as negative controls alongside the two *Psg22* shRNA vectors being tested.

LacZ transfections were performed using Lipofectamine 2000 (Invitrogen, 11668-019), and OptiMEM (Invitrogen, 31985-062) as per manufacturers instructions. Jar cells were seeded at a density of 5 x 10^4 cells/ml in 24 well plates. An empty LacZ vector was used as negative control for LacZ expression and a pCMV-SPORT-β-Gal construct (Life Technologies, 10586-014), was used as a positive control for LacZ expression in Jar cells. A Sprouty3 promoter LacZ construct was also used as a positive control. Three *Psg* promoter LacZ constructs were tested, *Psg20*, *Psg22* and *Psg23*.

Retinoic acid treatment of undifferentiated TSC cell lines (TS-EXE and TS-GFP) was performed using 5 µM ATRA (Sigma, R2625-100MG) and 5 µM 9-cis RA (Sigma, R4643-1MG) solubilised in 95% ethanol (EtOH), in 70FCM+F4H medium to induce differentiation to TGC. 5 µM EtOH was used as vehicle for control treatments. TSC seeded to a density of 5 x 10^4 in 24 well plates were incubated with ATRA or 9-cis RA for 24 and 48 hours in a humidified 5% CO_2 incubator at 37°C. Cells were harvested for RNA extraction and cDNA synthesis. For Psg22 purified protein treatments, RAW-246.7 and THP-1 cells around 90% confluence were incubated with 10 µg/ml Psg22 Long and 10 µg/ml Psg22 Short protein isoforms for 24 hours and then cell culture supernatant was harvested for ELISA. A Strep-His peptide (WSHPQFEKLEHHHHHHHH) (Eurogentec, Belgium) was used as a control for the Strep-His tag introduced to the C-terminus of the proteins expressed from the pQE-Trisystem-His-Strep-1 expression vector. This ensured that this introduced tag did not induce TGFβ1 expression in these cell lines.

2.3.4 PAC screen

A P1-Artificial Chromosome (PAC) 129/Sv RPCI.21 library was screened to obtain a *Psg23* positive clone, *Psg23* was chosen as it is positioned in the centre of the major Psg

cluster. The RPCI-21 PAC Library has been constructed with female 129S6/SvEvTac mouse spleen genomic DNA (partially *MboI* digested) and was cloned between the *BamHI* sites of the pPAC4 vector [248]. The average insert size is 147 Kbp. The library consists of approximately 128,899 clones in 336 microtitre plates. The plate numbers run from 337 to 672. The PAC library has been gridded onto 22x22 cm positively charged nylon filters for hybridization screening purposes. Each filter contains 36,864 colonies which represents 18,432 independent clones spotted in duplicate in a 4x4 clone array. Seven filters cover the whole library. This provides a 6-9 fold coverage of the mouse genome.

A probe approximately 2 kb upstream of *Psg23* (879 bp) was amplified from murine 129/Sv genomic DNA using primers: *Psg23* Probe F: 5'-TCCTGTCCCCACTAACCTTG-3' and *Psg23* Probe R: 5'-TGACAACCCCACACAAGAAA-3'. Amplified DNA was purified and cloned into pGEM-T Easy Vector (Promega, USA, A1360). Positive clones were sequenced and the 879 bp probe was removed from its vector backbone using *SalI* and *NcoI*. The *Psg23* probe was radiolabelled (α-P^{32}) dCTP (3000 Ci/mmole; Amersham) and the library was screened by hybridising P^{32}-radiolabeled *Psg23* probes to the library filters using Southern Blotting Hybridisation described elsewhere [248]. The blots were washed with 0.5X SSC and 0.1% SDS at 65°C and exposed to Kodak x-ray film (Kodak, USA) overnight at -80°C and results were analysed using the online clone identification protocol. Positive PAC clones were purchased from BACPAC resources at the Children's Hospital Oakland Research Institute (C.H.O.R.I, USA). Positive PAC clones were cultured, plasmid DNA was prepped using the Qiagen Large Construct kit (Qiagen, UK, 12462), as per manufacturers instructions, and characterised using gene specific primers. The primers used in the PAC characterisation are listed in Table 2.1. To characterise the PAC clone further, End Sequencing of the PAC was performed. Purified PAC DNA was sent to GATC (GATC Biotech, UK) to be sequenced using the T7 and SP6 promoter sequencing primers located on the pPAC4 plasmid and were provided by the company.

Table 2.1: PAC characterisation Primers

Primer	Sequence
Psg23 Upstream F	5′-TCCTGTCCCCACTAACCTTG -3′
Psg23 Upstream R	5′-TGACAACCCCACACAAGAAA-3′
Psg23 Downstream F	5′-TGGCAATGAGGAAATCAACAC -3′
Psg23 Downstream R	5′-GAGGGAGGAAAGAAGTCAGAGA -3′
Psg25 Specific F	5′-ACCCTCCACACACTGCTCTGCT-3′
Psg25 Specific R	5′-AGCAAACAAGGACACATGACACCA-3′
Psg27 Specific F	5′-CCATCCTGCCTGGTGCCTGC-3′
Psg27 Specific R	5′-CTCTCCCAGGGGTGGCCCTC-3′
Psg23 Specific F	5′-AGGGAGACCCACACTCACAC-3′
Psg23 Specific R	5′-AGGTAGTCCATGCCAGCAGT-3′
Psg21 Specific F	5′-GTCACATGACCCTGCCTTTT-3′
Psg21 Specific R	5′-GCAGAGGGGACCAAATTACA-3′
Psg20 Specific R	5′-GGAGTCAGCAGGTGTCAGCCC-3′
Psg20 Specific F	5′-TGAGCTGTGGGTGGTGGGGT-3′

2.3.5 Polymerase Chain Reaction

All Polymerase Chain Reactions (PCR) were performed using standard molecular techniques, using either Finnzymes Phusion Hot Start High Fidelity DNA Polymerase (ThermoScientific, F-530S) or Finnzymes Phusion 2 Hot Start High Fidelity DNA Polymerase (ThermoScientific, F-549S). PCR reactions were amplified using a G-Storm thermocycler (G-Storm, UK) in 50 µl reaction volumes. Reactions contained 10 µl of 5x GC Buffer, 1.2 µl of dNTPs (NEB, N0447L), 1.5 µl Forward primer, 1.5 µl Reverse primer, 2 µl DMSO, x µl template, 0.5 µl Phusion DNA Polymerase, made up to 50 µl with ddH_2O. Cycling conditions were 98°C for 3 minutes, 98°C for 30 seconds, x°C annealing for 40 seconds, 72°C for 1 or 2 minutes, 72°C for 10 minutes final extension. All PCR products were resolved on an agarose gel, composed of agarose and Tris-borate EDTA (TBE) buffer, using gel electrophoresis at 90 V for 50 minutes. PCR products were visualised using the UV Gel-doc system.

2.3.6 RNA extraction

Cells were lysed and RNA extracted at room temperature using TRI Reagent (Sigma, 93289-100ML). Phase separation was achieved by addition of chloroform, mixing vigorously and centrifugation at 12000xg for 15 minutes at 4°C. Nucleic acids present

in the upper aqueous phase were removed to a fresh 1.5 ml centrifuge tube and RNA was precipitated using ice-cold isopropanol. Nucleic acids were incubated at room temperature for 10 minutes. RNA was harvested by centrifugation at 1200x g for 5 minutes at 4°C. Supernatant was carefully removed and pellets were dried for 10 minutes at room temperature before resuspension in RNase-free ddH_2O. Nucleic acid concentration and purity was determined by spectrophotometry at 260 nm and 260/280 nm respectively.

2.3.7 Reverse Transcriptase Polymerase Chain Reaction

Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) was used to determine expression of transcripts in a variety of cell lines and tissue types. First strand cDNA was synthesised using 1 µg total RNA in a 20 µl reaction using random hexamer priming and the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, UK) as per protocol. RT-PCR was performed using either Finnzymes Phusion Hot Start High Fidelity DNA Polymerase (ThermoScientific, F-530S) or Finnzymes Phusion 2 Hot Start High Fidelity DNA Polymerase (ThermoScientific, F-549S). RT-PCR reactions were amplified using a G-Storm Thermocycler (G-Storm, UK) in 50 µl reaction volumes. Reactions contained 10 µl of 5x GC Buffer, 1.2 µl of dNTPs (NEB, N0447L), 1.5 µl Forward primer, 1.5 µl Reverse primer, 2 µl DMSO, x µl template, 0.5 µl Phusion DNA Polymerase, made up to 50 µl with ddH_2O. Cycling conditions were 98°C for 3 minutes, 98°C for 30 seconds, x°C annealing for 40 seconds, 72°C for 1 or 2 minutes, 72°C for 10 minutes final extension. Annealing temperatures were specific for each primer set, and an RT-PCR gradient protocol was employed to determine the optimal annealing temperature for each primer set. RT-PCR using gene specific primers for three marker genes of differentiation, was used to confirm whether TSC had differentiated correctly into TGC. Marker genes used to determine correct differentiation were: *Eomes*, a trophoblast stem cell marker, *TpbpA*, a SpT marker, and *Prolactin2 (Pl2)*, which is a TGC specific marker gene. Primers used are described in [44], and are listed in Table 2.2.

For identification of *PSG* transcript relative frequency in a variety of cell types

Table 2.2: Differentiation Marker Primers

Primer	Sequence
Eomes F	5′-TGATCATCACCAAACAGGGC-3′
Eomes R	5′-ACTGTGTCTCTGAGAAGGTG-3′
Pl-2 F	5′-TCCTTCTCTGGGGCACTCCTGTT-3′
Pl-2 R	5′-CCATGAAGGCTTTTGAAGCAAGATCA-3′
TpbpA F	5′-TGAAGAGCTGAACCACTGGA-3′
TpbpA R	5′-CAGGCAGTTCATATGTTGGG-3′

and tissues, expression surveys were performed using cloning and sequencing of RT-PCR products. Primer sets that amplify all known murine *Psgs* were designed in Wynne et al, 2006 [49]. A degenerative primer set: PsgF and PsgR, which amplifies all known murine *Psg* were designed, although *Psg22* and *Psg25* are of identical sequence in the amplicon generated by these primers. An amplicon of 124 bp was generated which was confirmed by gel electrophoresis on a 1.5% agarose gel. In order to distinguish between *Psg22* and *Psg25* and to ensure that there was no preferential amplification of any particular Psg, the above experiment was repeated using the primer set Psg-all2: Psgall2F and Psgall2R. An amplicon of 176 bp was generated which was confirmed by gel electrophoresis on a 1.5% agarose gel. For human tissue samples, two primer sets that amplify all known human *PSGs* were designed - PSGV4 and PSGV5. As above, two primer sets were designed to ensure that there was no preferential amplification of any particular *PSG*. Primer sequences are listed in Table 2.3. Amplicons were gel extracted using a Qiagen Gel Extraction Kit (Qiagen UK) and blunt cloned into the multiple cloning site (MCS) of the pSTBlue-1 cloning vector and transformed into NovaBlue Singles competent cells (Novagen, UK). Colonies were picked and cultured overnight in LB containing ampicillin at 50 µg/ml and plasmid DNA was extracted using a Qiagen spin mini-prep kit (Qiagen, UK). 10-20 individual recombinant clones containing the inserts of correct size from each amplification were sequenced (GATC Biotech, Germany).

2.3.8 *BY564540* antisense transcript characterisation

Once I had identified the *BY564540* antisense transcript as an interesting putative enhancer element by bioinformatical methods, it was necessary to discern whether

Table 2.3: PSG expression survey primers

Primer	Sequence
PSGF	5′-TYCAYCCDKTGGHTCTTCAAYA -3′
PSGR	5′-CACAYYGRTAMTYTCCASCATC-3′
Psg-All2F	5′-GTGTTGACAATCTGCCAGAGAATCTT -3′
Psg-All2R	5′-CTCCTGGGTGACATTTTGGATC -3′
Human PSG V4 F	5′-AGAGACCATGGGAACCCTCT-3′
Human PSG V4 R	5′-ATTCTGGATCAGCAGGGATG-3′
Human PSG V5 F	5′-AGCAGGGATGCATTGGAATA-3′
Human PSG V5 R	5′-ACAGCGCATCAAATGGAAG-3′

this EST was expressed and if so, to map the length of this antisense transcript. The original BY564540 EST and BLAST result sequences were used to design EST specific primers to investigate if the BY564540 EST and its three BLAST results were expressed in TSC and TGC. Primers used can be seen in Table2.4. Once it had been established that the BY564540 EST was expressed, an investigation concerning the length of the *BY564540* antisense transcript utilising RT-PCR primer walking was undertaken. RT-PCR was performed as described above. Primers used in the primer walking of the *BY564540* antisense transcript are shown in Table 2.5 and primers used in the primer walking of the *BLAST 1* antisense transcript are listed in Table 2.6. Primers were designed to amplify the antisense cDNA transcripts in a direction specific manner.

Table 2.4: *BY564540* EST and BLAST result expression primers

Primer	**Sequence**	**Product**	**Tm**
BY564540 Internal EST F *BY564540* Internal EST R	5′-AGATCCCAAGACTGCAGGAA-3′ 5′-GGCCCTCATCATAAGCACAT-3′	170 bp	57°C
BY564540 EST BLAST1 F *BY564540* EST BLAST1 R	5′-TCCCAAGACTGAACGTACTAT-3′ 5′-TTTTTTGGGCCTGAGAATCT-3′	137 bp	58°C
BY564540 EST BLAST2 F *BY564540* EST BLAST2 R	5′-TCCCAAGACTGCAGGAACTAC-3′ 5′-ATCCTTGAACCTGAGAATCT-3′	137 bp	57°C
BY564540 EST BLAST3 F *BY564540* EST BLAST3 R	5′-TCCCAAAACTGCATTCATTAA-3′ 5′-CTCCCTGGGTCCAAAAATCT-3′	137 bp	57°C

2.3.9 Quantitative Real Time Polymerase Chain Reaction

Quantitative Real Time Polymerase Chain Reaction (qRT-PCR) was performed as per [49], using the ABI PRISM 7900 sequence detection system (SDS) and the SYBR GREEN qPCR kit (Applied Biosystems, Foster City, CA, USA). RNA was extracted

Table 2.5: *BY564540* transcript characterisation primers

Primer	Sequence	Product	Tm
BY564540 Internal EST F *BY564540* Internal EST R	5′-AGATCCCAAGACTGCAGGAA-3′ 5′-GGCCCTCATCATAAGCACAT-3′	170 bp	57°C
BY564540 Internal EST F *BY564540* AS3 F	5′-AGATCCCAAGACTGCAGGAA-3′ 5′-TGCAAACAGTTATGGGGGAC-3′	669 bp	59°C
BY564540 Internal EST R *BY564540* AS3 R	5′-GGCCCTCATCATAAGCACAT-3; 5′-AGCGCCCTGTCTGGTTCCCT-3′	247 bp	58°C
BY564540 Internal EST R *BY564540* 4 R	5′-GGCCCTCATCATAAGCACAT-3; 5′-ATCCTACCAGTGGCTCTCAT-3′	270 bp	62°C
BY564540 Internal EST R *BY564540* 5 R	5′-GGCCCTCATCATAAGCACAT-3; 5′-CAGAAGGAGATGCCCAGTGA-3′	293 bp	59°C
BY564540 Internal EST R *BY564540* 6 R	5′-GGCCCTCATCATAAGCACAT-3; 5′-AAGTCTCATAAGCATTCAGAACA-3′	367 bp	60°C
BY564540 Internal EST R *BY564540* 7 R	5′-GGCCCTCATCATAAGCACAT-3; 5′-ACCATTGCCTGAAGGAGAGGA-3′	473 bp	57°C
BY564540 Internal EST R *BY564540* 8 R	5′-GGCCCTCATCATAAGCACAT-3; 5′-TGGATACTTGGCTGGAGACAGA-3′	681 bp	58°C
BY564540 Internal EST R *BY564540* 9 R	5′-GGCCCTCATCATAAGCACAT-3; 5′-GTAACCAAGTGATAGAGGACAAGGA-3′	1015 bp	58°C
BY564540 Internal EST R *BY564540* 10 R	5′-GGCCCTCATCATAAGCACAT-3; 5′-AGGGGAACATCAGCAGGTCA-3′	1436 bp	56°C
BY564540 A1 F *BY564540* 11 R	5′-TGACTGGGACTTGTTTACCTGAT-3; 5′-AGGAAGGCATGAGCAGATGA-3′	682 bp	58°C
BY564540 A2 F *BY564540* 11 R	5′-AAGCGTCGGATGAACTGACAA-3; 5′-AGGAAGGCATGAGCAGATGA-3′	787 bp	59°C
BY564540 A2 F *BY564540* 12 R	5′-AAGCGTCGGATGAACTGACAA-3; 5′-GCAGTTCAGGAGAGCAGAGCA-3′	918 bp	60°C
BY564540 A3 F *BY564540* 13 R	5′-TGTTGAACCCCCTGCTGTAG-3; 5′-TGGAGACAGACAGTGTGCTTCA-3′	772 bp	57°C
BY564540 A3 F *BY564540* 14 R	5′-TGTTGAACCCCCTGCTGTAG-3; 5′-TGCTCAGTCACTTCCACTCTCA-3′	1829 bp	56°C
BY564540 A3 F *BY564540* 15 R	5′-TGTTGAACCCCCTGCTGTAG-3; 5′-TCAGAGGACTTTGGGCTTCT-3′	6229 bp	60°C
BY564540 A3 F *BY564540* 16 R	5′-TGTTGAACCCCCTGCTGTAG-3; 5′-TGCTCTGTGGAATCCTCTACTCA-3′	7131 bp	59°C

from cell lines and tissues as previously described. First strand cDNA was synthesised using 1 µg total RNA in a 20 µl reaction using random hexamer priming and the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, UK, 4368814). The SYBR GREEN PCR master mix consists of Amplitaq Gold DNA polymerase, optimised PCR buffer, 25 mM $MgCl_2$, dNTP mix and AmpErase Uracil N-Glycosylase (UNG). All qRT-PCR reactions were performed in MicroAmp® Optical 384-Well Reaction Plates (Life Technologies, 4343370). PCR amplifications were performed in a total volume of 10 µl in triplicate wells. The following PCR protocol was used for all qRT-PCR reactions: denaturation program (95°C for 10 min), amplification and quantification program repeated for 40 cycles (95°C for 15 s, 58°C for 30 s,

Table 2.6: *BY564540 BLAST1* transcript characterisation primers

Primer	Sequence	Product	Tm
BY564540 EST BLAST1 F *BY564540* EST BLAST1 R	5′-TCCCAAGACTGAACGTACTAT-3′ 5′-TTTTTTGGGCCTGAGAATCT-3′	137 bp	56°C
BY564540 BLAST1 3.2 F *BY564540* BLAST1 3.3 R	5′-TTGGTATCTCAACAGCATCTTAATA-3′ 5′-TGAGACCCAGAAGGAGATGC-3′	863 bp	60°C
BY564540 EST BLAST1 F *BY564540* BLAST1 3.2 F	5′-TCCCAAGACTGAACGTACTAT-3′ 5′-TTGGTATCTCAACAGCATCTTAATA-3′	730 bp	60°C
BY564540 EST BLAST1 R *BY564540* BLAST1 3.3 R	5′-TTTTTTGGGCCTGAGAATCT-3′ 5′-TGAGACCCAGAAGGAGATGC-3′	270 bp	60°C
BY564540 EST BLAST1 R *BY564540* BLAST1 3.6 R	5′-TTTTTTGGGCCTGAGAATCT-3′ 5′-TGGTTCACAGACACCTGAGAA-3′	682 bp	60°C
BY564540 EST BLAST1 R *BY564540* BLAST1 3.7 R	5′-TTTTTTGGGCCTGAGAATCT-3′ 5′-TTCATTAAGACTGACTCCAAGA-3′	1152 bp	60°C
BY564540 EST BLAST1 R *BY564540* BLAST1 3.8 R	5′-TTTTTTGGGCCTGAGAATCT-3′ 5′-TAAGGTTATTTCTCTTTGGTCC-3′	1611 bp	60°C
BY564540 EST BLAST1 R *BY564540* BLAST1 3.9 R	5′-TTTTTTGGGCCTGAGAATCT-3′ 5′-TTTCACTCTTCTAAGTTCTCATAA-3′	2248 bp	60°C
BY564540 EST BLAST1 R *BY564540* BLAST1 4.0 R	5′-TTTTTTGGGCCTGAGAATCT-3′ 5′-CAGAAGCAGTTTAGGAGAGCAGA-3′	2600 bp	60°C
BY564540 BLAST1 4.0 F *BY564540* BLAST1 4.1 R	5′-TTTAGTCCATGACTTGCCAGG-3′ 5′-CACCCTTTCATCCCCAGAGTA-3′	700 bp	60°C
BY564540 BLAST1 4.2 F *BY564540* BLAST1 4.2 R	5′-TTTTCCTGGTTCAAGGGTGT-3′ 5′-AGGGAATTTGTAGGGACCAGA-3′	764 bp	60°C
BY564540 BLAST1 4.2 F *BY564540* BLAST1 4.3 R	5′-TTTTCCTGGTTCAAGGGTGT-3′ 5′-TTAACGCTCACATTGCTGTCTA-3′	1187 bp	60°C

60°C for 1 minute with a single fluorescence measurement), melting curve program (60°C – 95°C with a heating rate of 1°C per 30 s and a continuous fluorescence measurement). Thereafter, PCR products were identified by generating a melting curve, which was also used to assess the occurrence of putative PCR artefacts (primer-dimers) or non-specific PCR products. Normalisation of expression levels to the housekeeping gene, hypoxanthine-guanine phosphoribosyl transferase (*Hprt*), was used to avoid discrepancies caused by variations in input RNA or in reverse transcription efficiencies. Results were described as mean *Psg* expression relative to mean *Hprt* expression. Primers used for qRT-PCR reactions are shown in Table 2.7. Three biological replicates of each cell line were evaluated, using three technical qRT-PCR replicates.

2.3.10 Vector construction

A number of vectors used in this work were constructed as follows: *Psg22* short-hairpin RNA (shRNA) vectors were constructed as described in [249, 250]. The shRNA

Table 2.7: Quantitative Real-Time PCR primers

Primer	Sequence
Psg19 QRT F	5′-TCCAGTGCCACCACATGCTGTC-3′
Psg19 QRT R	5′-TGCACGGCCACTGATGATAGACTCT-3′
Psg21 QRT F	5′-AAACTGTGAATGGATTTCGGG-3′
Psg21 QRT R	5′-TGGAAGGAGGGAATTGGGTA-3′
Psg22 QRT F	5′-CGCATGGCCAGTTGGCCATT-3′
Psg22 QRT R	5′-AAAGCGGGGGAAATAGTTGTAGTA-3′
Psg23 QRT F	5′-GAGCCTGTCCCCGTCAAAGTGT-3′
Psg23 QRT R	5′-GAAATGCCTCTGCCCTGCTATAGT-3′
Hprt QRT F	5′-CTATAAGTTCTTTGCTGACCTGCT-3′
Hprt QRT R	5′-ATCATCTCCACCAATAACTTTTATGT-3′

oligonucleotide target sequences were designed using the PSICOOLIGIOMAKER1.5 software programme. *Psg22* coding sequence (CDS) was used as input sequence for the template. This programme predicts all the potential 19-mer oligonucleotide target sequences, and returns the sense and antisense oligonucleotides (5′ to 3′ orientation) required for gene silencing. These target sequences are listed in Table 2.8. The murine U6 promoter sequences (F 5′-TGTGCTCGCTTCGGCAGCACATATACT-3′ and R 5′-AGTATATGTGCTGCCGAAGCGAGCACA-3′) were incorporated before the shRNA target sequences to stabilize the shRNA and possesses downstream restriction sites (*HpaI* and *XhoI*) to allow the efficient introduction of oligonucleotides encoding shRNAs into the pSicoR-GFP vector. The CD8 oligonucleotide stem loops (F 5′-TTCAAGAGA-3′ and R 5′-TCTCTTGAA-3′) were used as described in [249]. Oligonucleotides were composed of U6 promoter sequence, CD8 stem loop, followed by the *Psg22* shRNA target sequence. Oligonucleotide target sequences were aligned against the mouse genome using the BLAST programme to ensure that these target sequences were *Psg22* specific. Two target oligonucleotides predicted to result in effective short-hairpin formation and gene silencing were picked, named *Psg22*shRNA 1 and *Psg22*shRNA 2. These target oligonucleotides targeted both splice variants of *Psg22*. 5′ phosphorylated oligonucleotides were purchased from MWG (MWG Eurofins, Germany). To construct the *Psg22* shRNA vectors, each oligonucleotide pair (Sense and Antisense) were initially annealed. 1 µl sense oligo (100 µM) and 1 µl antisense oligo (100 µM) were annealed in 25 µl 2x annealing buffer (200 mM potassium acetate, 60 mM HEPES-KOH pH 4, 4 mm Mg-acetate) for 4 minutes at 95°C,

10 minutes at 70 °C and then the reaction mix was slowly cooled to 4°C. Annealing of oligonucleotides was confirmed by gel electrophoresis in a 2% agarose gel. pSicoR purified plasmid was digested with *HpaI* and *XhoI* in parallel to oligonucleotide annealing. Correctly annealed oligonucleotides were then ligated into the purified digested pSicoR vector for three hours at room temperature using T4 ligase. 2 µl of each ligation reaction, *Psg22* shRNA 1-pSicoR and *Psg22* shRNA 2-pSicoR, were transformed as per protocol into Novegen competent cells. Positive clones were obtained, and sent to GATC for sequencing. Sequencing was performed using the *Psg22* shRNA sequencing primer: 5'-TGCAGGGGAAAGAATAGTAGAC-3'. Positive sequenced clones were then cultured and purified using the Endofree Plasmid Maxi Kit (Qiagen, 12163) as per protocol and stored at -20°C.

To assess the promoter activity of *Psg* promoters, *Psg* promoter LacZ reporter vectors were constructed as follows. *Psg* promoter regions, spanning a region 2 kb upstream of the transational start site (ATG), of *Psg20, Psg22, and Psg23* were amplified using primers with incorporated *NotI* restriction sites for ease of cloning. These primers were: *Psg* 2 kb Promoter F: 5'-ATAAGAATGCGGCCGCTTTGTGGTGTTGAACCCCCT-3' and the *Psg* Promoter R: 5'-ATAAGAATGCGGCCGCATCTCTTCTCACTGTACTGGCCTTT-3'. Psg promoter sequences were amplified from PAC3 purified DNA as described in PCR protocols above. Annealing temperature used was 68°C. The PCR products were digested with *NotI* for three hours at 37°C and purified using the Qiagen PCR purification kit (Qiagen, 28104) as per protocol. A LacZ reporter vector was digested with *NotI* for three hours at 37°C and gel extracted using the Qiagen Gel extraction kit (Qiagen, 28704) as per protocol. Purified digested PCR products were ligated into purified digested LacZ vector using T4 ligase, at 16°C for 10 hours. Ligations were transformed into Novegen competent cells. Positive clones were obtained, and sent to GATC for sequencing using the T7 promoter primers supplied by the company. Positive sequenced clones were then cultured and purified using the Endofree Plasmid Maxi Kit (Qiagen, 12163) as per protocol and stored at -20°C.

Table 2.8: *Psg22*shRNA 1 and *Psg22*shRNA 2 target oligonucleotide sequences

Primer	Sequence
*Psg22*shRNA 1 Sense	5′-GAAGAGAGATATTGTTCAT-3′
*Psg22*shRNA 1 Antisense	5′-ATGAACAATATCTCTCTTC-3′
*Psg22*shRNA 2 Sense	5′-GGACAGCACAGTTCGAATA-3′
*Psg22*shRNA 2 Antisense	5′-TATTCGAACTGTGCTGTCC-3′

2.3.11 Quantification of splice variants and antisense transcripts

Identification of an alternative splice variant of murine *Psg22* led to the investigation of the expression of this variant relative to the full length *Psg22* transcript expression in a variety of trophoblastic cell lines and tissues. Relative splice variant transcript quantification was performed as described elsewhere [251, 252], employing a dual insert plasmid containing specific regions of both transcripts to construct a standard curve for qRT-PCR analysis. Using E10 dissected TGC cDNA as template, transcript specific primers (*Psg22* Variant F and R) were designed to amplify a 608 bp region from the full length transcript and a 248 bp region from the truncated splice variant. Each RT-PCR amplicon was then individually gel extracted using the Qiagen gel extraction kit (Qiagen, 28704) and ligated into pSTblue1, using T4 ligase (Novagen perfectly blunt cloning kit, Merck, 70182-3) into the *Eco*RV restriction site in the MCS. Positive clones were picked and grown overnight in 5 ml LB with carbenicillin (50 µg/ml) at 37°C in shaking incubator. These overnight cultures were then miniprepped using the Qiagen Minispin kit (Qiagen, 28704) as per protocol. Plasmid DNA was then sent to GATC (GATC Biotech, Germany), to verify correct sequences had been inserted. These regions were then excised from the pSTBlue1 plasmids using restriction endonucleases *EcoRI* for the *Psg22* Long fragment; and *KpnI* and *XhoI* for the *Psg22* Short fragment. These products were gel extracted using the Qiagen gel extraction kit (Qiagen, 28704) and ligated sequentially into the MCS of pBluescript SK+ (Agilent Technologies, UK, 212205) plasmid using the same restriction endonucleases used to excise the fragments from pSTBlue1. Positive clones were picked and grown overnight in 5 ml LB with carbenicillin (50 µg/ml) at 37°C in shaking incubator. These overnight cultures were then miniprepped using the Qiagen Minispin kit (Qiagen, 28704) as per protocol. The dual insert pBluescript SK+ plasmid was then sent to GATC (GATC

Biotech, Germany), to verify that both sequences had been correctly inserted. Once both inserts had been correctly cloned and sequence verified, a standard curve was constructed using serial dilutions of the template plasmid. Two standard curves are generated from the same serial dilutions, thus providing complete equality of both curves as described [251]. Correctly diluted dual insert vector standard curves were used to perform relative quantification of each transcript using qRT-PCR. qRT-PCR was performed as described in the qRT-PCR section above. Primers used in the cloning of the dual transcript vector (*Psg22* Variants primer set) and in the qRT-PCR reactions (*Psg22* Long and *Psg22* Short primer sets) are described in Table 2.9.

Identification of the *BY564540* EST antisense transcript led to the investigation of the expression of this transcript relative to the full length *Psg22* transcript expression in a variety of trophoblastic cell lines and tissues. Relative quantification of these transcripts was performed as described elsewhere [251], employing a dual insert plasmid containing specific regions of both transcripts to construct a standard curve for qRT-PCR analysis. Using E10 dissected TGC cDNA as template, transcript specific primers, EST *BY564540* (EST7R and IESTF primer set) and *Psg22* (*Psg22* Variants primer set), were used to amplify a 473 bp region of the BY564540 EST transcript and a 608 bp region of the *Psg22* transcript. Each RT-PCR amplicon was then individually gel extracted using the Qiagen gel extraction kit (Qiagen, 28704) and ligated into pSTblue1 using T4 ligase (Novagen perfectly blunt cloning kit, Merck, 70182-3) into the *Eco*RV restriction site in the MCS. Positive clones were picked and grown overnight in 5 ml LB with carbenicillin (50 µg/ml) at 37°C in shaking incubator. These overnight cultures were then miniprepped using the Qiagen Minispin kit (Qiagen, 28704) as per protocol. Plasmid DNA was then sent to GATC (GATC Biotech, Germany), to verify correct sequences had been inserted. These regions were then excised from the pSTBlue1 plasmids using restriction endonucleases *EcoRI* for the *Psg22* Long fragment; and *KpnI* and *XhoI* for the *BY564540* EST antisense transcript fragment. These products were gel extracted using the Qiagen gel extraction kit (Qiagen, 28704) and ligated sequentially into the MCS of pBluescript SK+ (Agilent Technologies, UK, 212205) plasmid using the same

Table 2.9: Splice variant quantification primers

Primer	Sequence
Psg22 Variants F	5′-GGAGGTATCCTCTGAGCTTCTCA-3′
Psg22 Variants R	5′-TTCTGTGCCGAGCAATCTCAA-3′
Psg22 Long F	5′-TTCTGCTCACAGCCTCCCTCT-3′
Psg22 Long R	5′-ACCCCTCTATACCAGACAAAGACTCGAA-3′
Psg22 Short F	5′-TCTGCTCACAGCCTCTCTTTTCA-3′
Psg22 Short R	5′-TTGTACCAGAGAAGCGATTGAAGA-3′

Table 2.10: *Psg22* and *BY564540* antisense transcripts quantification primers

Primer	Sequence
Psg22 Variants F	5′-GGAGGTATCCTCTGAGCTTCTCA-3′
Psg22 Variants R	5′-TTCTGTGCCGAGCAATCTCAA-3′
BY564540 EST - IEST R	5′-GGCCCTCATCATAAGCACAT-3′
BY564540 EST - EST7R	5′-ACCATTGCCTGAAGGAGAGGA-3′
Psg22 QRT F	5′-CGCATGGCCAGTTGGCCATT-3′
Psg22 QRT R	5′-AAAGCGGGGGAAATAGTTGTAGTA-3′
BY564540 EST - IEST F	5′-AGATCCCAAGACTGCAGGAA-3′
BY564540 EST - IEST R	5′-GGCCCTCATCATAAGCACAT-3′

restriction endonucleases used to excise the fragments from pSTBlue1. Positive clones were picked and grown overnight in 5 ml LB with carbenicillin (50 µg/ml) at 37°C in shaking incubator. These overnight cultures were then miniprepped using the Qiagen Minispin kit (Qiagen, 28704) as per protocol. Dual insert pBluescript SK+ plasmid was then sent to GATC (GATC Biotech, Germany), to verify correct sequences had been inserted. Once both inserts had been correctly cloned and sequence verified, a standard curve was constructed using serial dilutions of the template plasmid as described [251]. Correct standard curves were used to perform relative quantification of each transcript using qRT-PCR and the standard curve created with the dual insert vector. qRT-PCR was performed as described above. Primers used in cloning (*Psg22* Variants and *BY564540* IEST R-EST7 R primer sets) and in the qRT-PCR reactions (*Psg22* Long and *BY564540* IESTFR primer sets) are described (Table 2.10.).

2.3.12 ELISA

For the ELISAs, cells were plated in triplicate wells for each treatment in 24 well plates and incubated in a 37°C humidified incubator with 5% CO_2. Raw246.7 cells and THP-1 cells were seeded at a density of 1 x 10^6 cells/ml per well. Cells were treated on

the following day in 300 µl of fresh media for 24 hours. Cells were also treated with recombinant PSG1 protein as positive control, and Strep-His peptide as a negative control. After treatments, the supernatants were collected and centrifuged at 3000 rpm for 5 minutes to remove cell debris. For TGFβ1 ELISA, supernatant was activated as per the manufacturer's instructions. Induction of TGFβ1 in human monocytic and murine macrophage cell lines by recombinant Psg22 proteins was quantified using the Human/ Mouse TGFβ1 ELISA Ready-SET-Go Kit (eBiosciences, 88-7449) as per manufacturers instructions. This ELISA is engineered for quantification of mouse or human TGFβ1 protein levels from supernatants from cell cultures. This ELISA has a sensitivity of 60 pg/ml.

2.3.13 β-Galactosidase Assay

The quantification of β-galactosidase activity from LacZ-reporter constructs was performed using the Pierce Mammalian β-Galactosidase Assay Kit (ThermoScientific, 75707). The Thermo Scientific Mammalian β-galactosidase Assay Kit provides a colorimetric method for lysing cultured mammalian cells and measuring β-galactosidase activity. This kit was used to quantify LacZ expression driven by *Psg*-promoter-LacZ constructs in transfected cell lines. *Psg*-promoter-LacZ vectors were constructed as described below. Empty LacZ vector was used as a negative control, and the pCMV-SPORT-βGal construct (Life technologies, 10586-014) was used as a positive control, as this construct drives LacZ expression through the strong cytomegalovirus (CMV) promoter. Jar cells were plated in 24 well plates at a density of $2x10^5$ cells/ml and cultured as described above. Cells were transfected using Lipofectamine2000 as described above and cultured for 24 hours post-transfection. The Mammalian β-Galactosidase Assay Kit was used as per manufacturers instructions. The absorbance was read at 405 nm every hour until the absorbance remained static using a Spectramax384 Plus Absorbance Microplate Reader (Molecular devices, USA).

Table 2.11: Chromatin accessibility primers

Primer	Sequence
Psg22 CA F	5′-CCCTTCCCAGAGCACTGAGGACACA-3′
Psg22 CA R	5′-AGCACTGACATGCCCCCAGAGAACA-3′
Psg23 CA F	5′-CCACGTCCAGGAGTCAGCAGATGTC-3′
Psg23 CA R	5′-GAGGGAGGAAAGAAGTCAGAGA-3′
BY564540 CA F	5′-GGGCCTGAGAATCTGGCTGCTGAAA-3′
BY564540 CA R	5′-TGTGCTCTCCATGCTGAGACCCAGA-3′
B1 CA F	5′-GGCCTGAGAATCTGGCTGCAGAAAC-3′
B1 CA R	5′-TGCTCTCCATGCTGAGACCCAGAAG-3′
BY564540 2kbUP CA F	5′-TTGAGCGTTCCTGGCTCTGAGTGTC-3′
BY564540 2kbUP CA R	5′-CCTGGGCCTCCTGCATCAGTTAAGA-3′
BY564540 2kbDWN CA F	5′-GCACCCCAACACATGCGAAAACCTA-3′
BY564540 2kbDWN CA R	5′-GTTTCCATCTCCAGCGTTGCCTCAC-3′
B1 2kbUP CA F	5′-GCCTTGACTTCCTGCAGGGCTACAC-3′
B1 2kbUP CA R	5′-CTCACTGGCCCATGTCTGGTGTCTC-3′
B1 2kbDWN CA F	5′-GCTGAGTATGCATCTCCCCCAGGTC-3′
B1 2kbDWN CA R	5′-CAGCCAAAGCCAAACCAGGAGACTG-3′
Gadph Control F	5′-CAGCTCCCCTCCCCCTATCAGTTCG-3′
Gadph Control R	5′-ACCAGGGAGGGCTGCAGTCCGTATT-3′
Rho Reference F	5′-AGGTCACTTTATAAGGGTCTGGGGG-3′
Rho Reference R	5′-AGTTGATGGGGAAGCCCAGCACGAT-3′

2.3.14 Chromatin Accessibility assay

Chromatin accessibility in specific genomic regions of the murine *Psg* locus was measured using the EpiQ Chromatin Accessibility Assay kit (BioRad, 172-5400) as per manufacturers instructions. TSC lines (TS-R26 and TS-GFP) and their differentiated TGC, MEFs, and 3T3 cells were grown as previously described. *In situ* nuclease digestion was performed, cells were lysed and qRT-PCR was performed using the Roche Lightcycler 480 system (Roche, UK, 05015243001) as per protocol. Primers used were designed according to the manufacturers instructions and using Primer3 software (http://frodo.wi.mit.edu/). Primer efficiency was calculated by the formation of a serial dilution standard curve and efficiency was analysed using the EpiQ Chromatin Kit Data Analysis Tool software. The murine Reference (Rhodopsin, *Rho*) and Control (Glyceraldehyde 3-phosphate dehydrogenase, *Gadph*) gene primers used were supplied with the kit. All primer sequences are described in Table 2.11. Percentage Chromatin Accessibility was then quantified using the EpiQ Chromatin Kit Data Analysis Tool software supplied with the kit as per protocol.

2.3.15 Polysome fractionation

This technique is a slight modification of previously reported methods [253]. This technique allows the fractional determination of a specific mRNA (*Psg22*) and whether this transcript is bound to ribosomes or exists as a free mRNA particle. This gives an estimation of the transcripts translational efficiency. In this technique free mRNAs and polysome-bound mRNAs are separated by the principle of sedimentation velocity in a sucrose gradient. Cycloheximide (C7698, Sigma) was used to immobilize ribosomes on mRNAs. While free mRNAs will not enter the gradient, the migration of ribosome-bound transcripts is directly proportional to their loading with ribosomes, due to increase in density of polysomes over free mRNAs [253]. E10 CD1 dissected TGC tissue (approx. 20 mg) was used as a sample. Sample processing involved pulverization of the tissue with a precooled mortar and pestle. This step requires maintaining the tissue frozen: the mortar is filled with liquid nitrogen with the pestle inside. Once cold, the tissue is added and then pulverized until a fine powder is obtained, adding more liquid nitrogen when necessary. This powder was then lysed using 1 ml of a NP40 lysis buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl, 15 mM $MgCl_2$, 20 mM DTT, 100 µg/ml cycloheximide, 0.5% Triton-X, 24 U/ml DNase, 20 U/ml Rnasin, 40 mM VRC, and 1% NP40). Nuclei were then removed by microcentrifuging at 12000 xg for 10 seconds at 4°C. Cytoplasmic extract was loaded onto 11 ml 10-60% sucrose gradients (10 and 60% m/v sucrose, 20 mM Tris-HCl pH7.5, 250 mM NaCl, 15 mM $MgCl_2$, 1 mM DTT, 100 µg/ml cycloheximide).

Sucrose gradients were made as described elsewhere [254]. Gradients were then run for three hours at 38,000x g at 4°C in a Beckman Coulter SW41Ti Swinging-bucket rotor in an Ultracentrifuge with no brake applied. After centrifugation, 40 x 300 µl fractions were collected carefully from the top and stored at -80°C. Total mRNA in each fraction was determined using A260/280 UV spectrometer. The 40 fractions were added to their neighbouring fraction to create 20 fractions, which were then used for RNA extraction. Fractions 20-40 are diluted with ultrapure H_2O to allow for dilution of concentrated sucrose. Each fraction was supplemented with 30 µl of 0.5 M EDTA (pH 5.1), 30 µl of 10% SDS (to allow dissociation of ribosomes), and 600 µl of

phenol-chloroform-isoamyl alcohol mixture acidic pH (4-5). Samples were vortexed and then the upper aqueous phase was placed into a new tube supplemented with 60 µl 3 M NaOAc pH5.1, 2 µl GlycoBlue, and 1 ml isopropanol. This was stored at -80°C overnight. Fractions were thawed and microcentrifuged at 12,000 xg for 15 minutes at 4°C. The pellets were then washed with 80% EtOH, the pellet was then dried the pellet and dissolve it in 50 ul H_2O. Purified RNA concentrations were then determined using UV Spectrometer at A260 nm, and stored at -80°C. RNA was then used in cDNA synthesis as described above and qRT-PCR was used to determine which sucrose fractions contained *Psg22* transcripts and the translational efficiency of *Psg22*. *Psg22* qRT-PCR primers and *Hprt* qRT-PCR primers were used to amplify transcripts of interest, primer sequences are listed (Table 2.7.).

2.3.16 Protein production

Both splice variant isoforms of *Psg22* were amplified by RT-PCR from E15 placental cDNA synthesised with Applied Biosystems High Capacity cDNA synthesis kit (Applied Biosystems, Life Technologies, UK, 4368814) incorporating the restriction sites *NcoI* and *PmlI*. Primers used were: *Psg22* ORF F: 5'-CATGCCATGGAGGTATCCTCTGAGCTTCTCAGCAATG-3' and *Psg22* ORF R: 5'-CACGTGCCTCATTCATCACAGTCAGCCTGACTGG-3'. This primer set amplified two transcripts, yielding products of 1425 bp and 1069 bp respectively. These products were gel extracted using the Qiagen gel extraction kit (Qiagen, 28704) and ligated into pSTblue1 using T4 ligase (Novagen perfectly blunt cloning kit, Merck, 70182-3) into the *Eco*RV restriction site in the MCS. Positive clones were picked and grown overnight in 5 ml LB with carbenicillin (50 µg/ml) at 37°C in shaking incubator. These overnight cultures were then miniprepped using the Qiagen Minispin kit (Qiagen, 28704) as per protocol. Plasmid DNA was then sent to Macrogen (Macrogen, The Netherlands) for sequencing. Positive sequences revealed that this RT-PCR product was indeed a splice variant of *Psg22*. Positive clones were cultured, miniprepped, and then purified plasmid was digested with restriction endonucleases *NcoI* (R0193S) and *PmlI* (R0532S) in NEBuffer 1 and BSA at 37°C for 2 hours. Digests were ran

through a 0.8% agarose gel, and the correct bands were excised and gel purified using the Qiagen gel extraction kit. Purified products were then ligated into *NcoI-PmlI* digested empty pQE-TriSystem-His-Strep1 expression vector (Qiagen, 32942). Ligations were performed at 16°C overnight in a G-storm thermocycler. Ligations were then transformed into NEB Turbo Competent cells (NEB, C2984H) as per protocol and transformation reactions were then plated onto Agar plates with 50 µg/ml carbenicillin and X-gal (70 µg/ml) and IPTG (80 µM). Plates were placed in 37°C incubator overnight. Positive colonies were picked and grown in 5 ml LB with carbenicillin (50 µg/ml) at 37°C in shaking incubator overnight. Overnight cultures were miniprepped using Qiagen Minispin kit (Qiagen, 28704) as per protocol. Plasmid DNA was digested with *NcoI* (R0193S) and *PmlI* (R0532S) in NEBuffer 1 and BSA at 37°C for two hours. Digest reactions were run through 0.8 agarose gel. Correct digestion patterns confirmed correctly cloned inserts. Positively digested clones were then sent to GATC (GATC Biotech, Germany) for sequencing using sequencing primers: *Psg22* Seq1 F: 5'-GTTATTGTGCTGTCTCAT-3' and *Psg22* Seq1 R: 5'-ATCGATCTCAGTGGTATTTGTG-3'. Positive sequencing data confirmed that both *Psg22* transcripts, *Psg22*-Long and *Psg22*-Short were cloned correctly in-frame into pQE-TriSystem-His-Strep1 expression vector.

To produce recombinant Psg22 protein isoforms, endotoxin-free plasmid DNA was purified from *Psg22* Long and Short pQE bacterial cultures using the Endofree Plasmid Maxi Kit (Qiagen, 12163). All subsequent steps were carried out using confirmed endotoxin-free reagents and tissue culture flasks. The DNA was transiently transfected into Freestyle™ 293-F cells (Life Technologies, K9000-01) using Freestyle™ MAX reagent Life Technologies, 16447750). The Freestyle™ 293-F cells were grown in suspension in Freestyle™ 293 Expression Medium (Life Technologies, 12338-001), by shaker culture, to a density of 1 x 10^6 cells/ml. The plasmid DNA was diluted in OptiPRO™ Serum Free Medium (Invitrogen, 12309-050) at a ratio of 1 µg DNA in 20 µl OptiPRO™ for every 1 ml of cells. Freestyle™ MAX reagent was also diluted in OptiPRO™ at the same ratio (1 µl Freestyle™ MAX reagent in 20 µl OptiPRO™ per ml of cells). The diluted DNA and Freestyle™ MAX reagent

were then combined, mixed gently and incubated at room temperature (RT) for 20 minutes. The mixture was added to the cell suspension and the cells were cultured for a further 72 hours. The culture was then centrifuged at 1,000 rpm for 5 min at RT to separate the protein-containing medium from the cells, and the medium was frozen in aliquots at -80°C. Recombinant Psg22 proteins were purified from cell culture medium by affinity chromatography using Qiagen Ni-NTA resin (Qiagen, 30210). Imidazole (Sigma-Aldrich; St. Louis, Missouri, I5513-25G) was added to the culture medium to a final concentration of 10 mM to reduce non-specific binding. Ni-NTA resin was added to the medium at a ratio of 1 ml resin suspension (corresponding to 0.5 ml resin bed volume) to 100 ml medium. The medium and resin were then batch bound overnight on a rotating wheel at 4°C. The medium and resin mix was then passed through a disposable polypropylene column (Pierce, Thermo Fisher Scientific; Ireland, 29924) and the resin was washed with wash buffer (500 mM NaCl, 20 mM NaH_2PO_4, pH 6, until the absorbance at 260 nm reduced to 0. Protein was then eluted from the column with increasing concentrations of imidazole in wash buffer, 4 x 1.5 ml 50 mM fractions, 5 x 1.5 ml 200 mM fractions, 4x 1.5 ml 300 mM fractions and 3 x 1.5 ml 500 mM fractions. Psg22 proteins were generally observed to elute in the five 200 mM fractions and the 300 mM fractions. These Psg22 containing fractions were then pooled and concentrated to a volume of 4 - 6 ml using a Millipore Amicon® Ultra Ultracel 10K centrifugal filter (Millipore, Ireland, UFC901024). The concentrate was then dialysed against three changes of 2 L of phosphate buffered saline (PBS) at 4°C. The protein was then further concentrated to a volume of 1 - 2 ml depending on the starting volume of culture medium. Purified recombinant protein was quantified by the Extinction Coefficient method, purity was checked by polyacrylamide gel electrophoresis using Coomaisse blue staining (Sigma, G1041), and tested for LPS contamination using Limulus Amebocyte Lysate QCL-1000 (Cambrex BioScience; Karlskoga, Sweden). Purified proteins were then aliquoted and frozen at -80°C.

2.3.17 Polyacrylamide Gel Electrophoresis SDS-PAGE and immunoblotting

Protein extracts were prepared by washing cells with PBS and lysing in lysis buffer (Tris HCl, pH 7.4, 150 mM NaCl, 1% NP40, and the protease inhibitors PMSF (1 mM), pepstatin (1 µM) and aprotinin (1.5 µg/ml). After incubation at 4°C for 20 minutes nuclear and cellular debris were removed by microcentrifugation at 14,000 rpm for 15 minutes at 4°C. Total protein was quantified using BCA Protein Assay Kit (Merck, 71285) according to manufacturer's protocol and lysate was stored at -80°C. For all Coomaisse stained gels and western blotting, protein preparations from HEK293 cell lysates and purfied proteins were resolved by sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis and transferred to a nitrocellulose membrane. Proteins were mixed with Laemmli sample buffer (10% glycerol, 2% SDS, 0.01% bromophenol blue, 5% 2-mercaptoethanol, 50 mM Tris, pH 6.8) and boiled at 95°C for 5 minutes prior to gel loading. Proteins were separated using the Bio-Rad Mini-Protean II gel electrophoresis system. Gels were resolved initially at 20 milliamps until the protein had passed through the stacking gel and then at 35 milliamps for approximately 1.5 hours until the dye front had reached the bottom of the gel. For Coomassie stained gels, the gel was removed from the electrophoresis apparatus and incubated in Coomassie blue stain (1 g/L Coomassie brilliant blue R-250, 25% 2-propanol, 10% acetic acid) for at least 1 hour before being destaining in a solution of 7% acetic acid and 25% EtOH for 2 hours. Gels were then imaged using the protein gel module of the Odyssey infrared scanning system (LI-COR).

For Western blotting, the gel was removed from the electrophoresis apparatus and pre-equilibrated in Transfer buffer (48 mM Trizima, 38 mM glycine, 0.037% SDS, 10% EtOH). The gel was then placed on top of a piece of nitrocellulose of the same size as the gel and sandwiched between three sheets of identically-sized filter paper that were pre-equilibrated in Transfer buffer. Bubbles were removed using a roller and proteins were transferred using the Bio-Rad Trans-Blot semi-dry transfer system at 18 V for 26 min. Transferred protein was confirmed with Ponceau Red staining and the membrane was then blocked in 5% non-fat dry milk (Marvel) in PBS containing 0.1%

Tween (PBS-T) for 1 hour at RT. Membranes were probed with either rabbit anti-His-Tag pAb diluted 1:1000 in 5% milk or rabbit anti-Psg22N1A mAb diluted 1:800 in 5% milk and PBS overlaid onto the membrane for at least 1 hour with rocking or overnight at 4°C. Following this, the membrane was washed (three 5 minute washes) with TBS-T (TBS supplemented with 0.1% Tween-20). Secondary antibody goat anti-rabbit IRDYE 680 (LI-COR) was then diluted 1:1000 in a 5% milk/TBS solution and overlaid onto the membrane for at least 1 hour with rocking. The membrane was washed again (three 5 min washes) with TBS-T and then a final wash in TBS. Membranes were then imaged using the membrane module on the Odyssey infrared scanning system (LI-COR). Protein molecular weight markers were purchased from New England Biolabs unless otherwise stated. Recombinant murine Psg17N1 and Psg22N1A proteins and murine anti-Psg antibodies were obtained as a gift from G. Dveksler. Recombinant Psg17N1 protein was produced as described previously in [164]. Recombinant Psg22N1A protein was produced as described in [165]. Rabbit polyclonal anti-Psg22N1A and rabbit polyclonal anti-Psg17N1 antibodies were generated by GenScript (USA), as described in [165]. These anti-Psg antibodies were tested by Western immunoblot for cross-Psg reactivity, and/or cross-species reactivity. Western immunoblots were carried out as described above using 2 µg of each purified recombinant protein, including mouse Psg22 Long and Short isoforms, Psg22N1A, and Psg17N1, human PSG1, PSG9 and BSA standard as samples.

2.3.18 Data and Statistical Analysis

All graphs were created using GraphPad Prism Software (GraphPad Software Inc, La Jolla, CA, USA). p values were determined by means of one way ANOVA and Bonferroni's multiple comparison post-test, with $p<0.05$ being deemed statistically significant. (n=) number of biological replicates.

Chapter 3

Results

3.1 Bioinformatics and Expression

3.1.1 Introduction

To improve our understanding of the rodent *PSG* multigene families, a complete map of the *PSG* loci in mouse and rat, is essential. In this chapter I investigated the structure and organisation of the rodent *PSG* loci using sequence data and bioinformatic techniques. Discerning the correct *PSG* locus structure will help in our understanding of *PSG* evolution and this complex multigene family's expansion. Using phylogenetic tree building software I constructed phylogenetic tree alignments of both murine and rat *PSG* families to discern if these species had orthologous relationships and whether they underwent similar family expansions. I found that the uncharacterised mouse *Psg31* and *Psg32* genes were incorrectly annotated as a pseudogene (LOC381852) and a hypothetical gene (*Psg-ps1*), respectively. RT-PCR products were cloned and sequenced to confirm expression of these genes in E15.5 murine placenta. I investigated further the expression patterns of *Psgs* in a variety of trophoblastic tissue lineages and TSC lines using cloning of RT-PCR products and qRT-PCR. PSG staining was detected in immunohistochemical sections of human gastrointestinal tract (GIT) (A. Houston & T. Moore, personal communication). I investigated whether *PSGs* were expressed in the murine and human GIT by RT-PCR and qRT-PCR methods. I found

that *Psg22* had an alternative splice transcript, and investigated the abundance of this transcript relative to the primary full lenght *Psg22* transcript in TSC, differentiated TGC, and a variety of trophoblastic tissues. As *Psg22* is the most abundant *Psg* transcript in the first half of pregnancy, I investigated whether this transcript was being translated efficiently using a polysome fractionation assay utilizing a sucrose gradient. These experiments have generated a set of expression data which will enhance our knowledge of this complex multigene family.

3.1.2 Reviewing the human and rodent *PSG* loci: Structure, organisation and orthology

Following initial investigations concerning the correct *PSG* locus organisation by McLellan et al, 2006 [148], I wanted to investigate if these predictions agree with the current build of the human and rodent genomes. Previous locus organisation predictions were based upon *PSG* sequence-specific oligonucleotide probing of YAC clones and large cosmids, genome walking and previous genome builds, and since then the genome assemblies have been resequenced, and better organised and annotated. Currently, the human genome build is the Genome Reference Consortium GRCh37 build. The current mouse genome build is the Genome Reference Consortium GRCm38 and the current rat build is the RGSC Rnor5.0 assembly. All of these genome assemblies were published in 2011 and are the most current genome assemblies for each of these species. The mouse *Psg* locus is on proximal chromosome 7, while the rat *PSG* locus is located on rat chromosome 1. Using existing sequence data from RefSeq libraries (NCBI), Ensembl and UCSC genome browsers, an accession table of all known mouse, rat and human PSG was compiled (Table 1.3). Using this data, all known murine *Psg* mRNA sequences from each of these databases, and using the BLAST program, every known mouse *Psg* was aligned to an approximately 2 Mb sequence taken from NCBI:M38:7:17566974 to 19627308 of mouse chromosome 7. Also using all known rodent *PSG* mRNA sequences from each of these databases, and using the BLAST program, every known rat *PSG* was aligned to an approximately 1.3 Mb sequence taken from NCBI: RGSC3.4:1:77301714 to 78604399 of rat chromosome

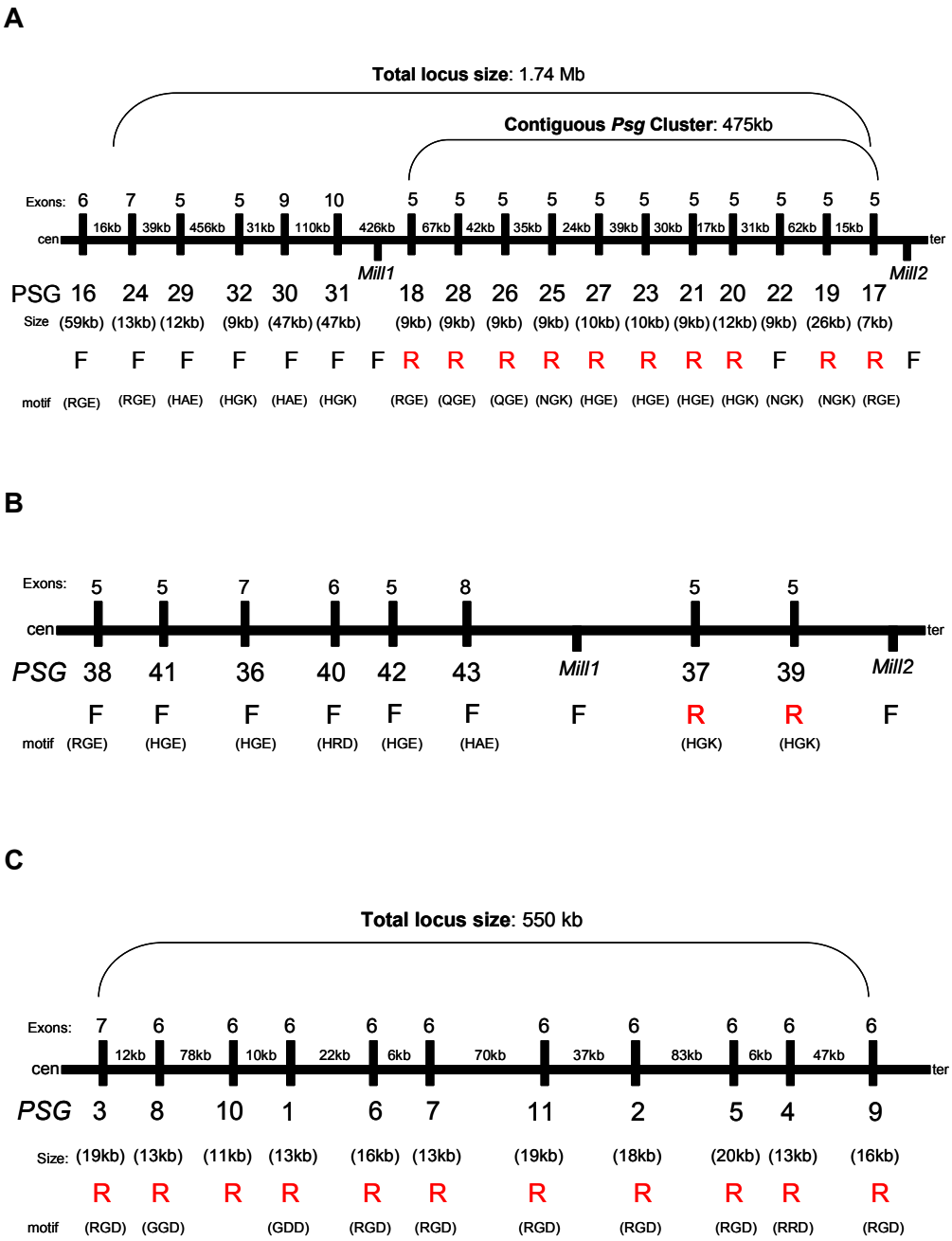


Figure 3.1: Rodent and human *PSG* loci structure and organisation. (A) Murine *Psg* locus on chromosome 7 including exon number, distance between *Psg* genes, genomic size, and strand orientation. (B) Rat *PSG* locus organisation on chromosome 1. The distance between rat *PSG* and their genomic size is omitted due to incompleteness of the rat genome. (C) Human *PSG* locus organisation on chromosome 19q13.2. F, gene in forward strand; R, gene in reverse strand.

1. From these data I produced updated maps of the rodent *PSG* loci showing gene length, distances between genes, and the orientation of each *PSG* in the loci. Rodent and human *PSG* loci maps are shown (Fig:3.1.). The identification of syntenic regions, which are blocks of genes or other markers demonstrating an evolutionary conserved order, and the quantification of evolutionary relatedness between genomes in terms of chromosomal rearrangements is one of the main research goals in comparative genomics [255]. With the advent of advanced sequencing technologies there has been continued growth of genomic sequence data from different species within public databases, and comparative mapping using bioinformatical tools has become increasingly important in the identification of functionally related genes within regions of interest across a range of species. Comparative genome mapping approaches are based on the sequence conservation between species and allow the data generated in model organisms such as the mouse and rat to be related to the human genome.

From our locus maps (Fig:3.1.A&B) we can see that the *PSG* loci of both the mouse and rat are quite similar in structure. Both loci contain a Major Histocompatibility Complex 1-like (MHC1-like) leukocyte (*Mill1* and *Mill2*) gene flanked *PSG* cluster, although this cluster of *PSG* genes in the rat has not undergone as big an expansion of family members as in the mouse *Psg* major cluster which contains 11 of the 17 murine *Psg*. Or the rat *PSG* family has undergone a contraction compared to the common ancestor. We can also see that all the murine *Psg* located in this *Mill1/2* flanked major *Psg* cluster are structurally similar, each containing 5 exons that contribute to three N domains and 1 A domain. The distances between murine *Psg* genes in the major *Psg* cluster are shorter in comparison to the murine *Psg* genes located outside the major cluster. There is approximately 450 kb between *Psg29* and *Psg32*, likewise, there is approximately 425 kb between *Psg31* and *Psg18* which is located in the major *Psg* cluster. *Psg16* has the longest murine *Psg* gene length of approximately 60 kb in comparison to the rest of the murine *Psg* famliy members which are on average about 10 kb long. The orientation of the rodent *PSG* genes that are flanked by the MHC1-like leukocyte 1 and 2 genes (*Mill1* and *Mill2*), are

located on the reverse DNA strand, with the exception of the mouse *Psg22* gene which, perhaps significantly, is the most abundant *Psg* transcript in the TGC in the first half of pregnancy [49]. This is evidence that *Psg22* has undergone an independent inversion event during the *Psg* family expansion. With *Psg22* located on the positive strand, and the remainder of the Mill1 and Mill2 flanked *Psgs* located on the negative strand, this gene specific inversion event, differentiates *Psg22* from the rest of the murine *Psg* family. This inverted orientation may help to explain how *Psg22* has increased expression relative to its family members. The human *PSG* locus also has varying distances between each *PSG* gene, some being only 6 kb apart, while others can be approximately 80 kb apart (Fig:3.1.C). It is of note that the human *PSGs* are smaller than mouse *Psgs*, and are between 9 kb and 20 kb long. All human *PSG* genes are located on the reverse strand, similar to the rodent loci.

To assess the homology between the rodent *PSGs*, phylogenetic analysis was performed using full length CDS sequences of both mouse and rat *PSGs*. Species specific PSG family trees were constructed to assess *PSG* homology in mouse, rat and human *PSG* families. These NJ trees (Fig:3.2.), and are constructed as previously described. The murine *Psgs* that are contained in the major *Psg* cluster are located on one major branch of the tree, while the *Psgs* located outside this major *Psg* cluster are branched together (Fig:3.2.A). Rat *PSG37* and *PSG39* branch together (Fig:3.2.B). This phlyogenetic analysis has also shown that there are orthologous relationships between certain mouse and rat *PSG* gene family members when an NJ tree of both rodent species is constructed (Fig:3.3.A). It was first suggested in McLellan *et al*, 2005, that these orthologous relationships existed, but these trees were constructed using incomplete *PSG* family sequences. Using the current rodent *PSG* CDS sequences and loci structure, neighbour-joined pairwise comparison phylogenetic trees were constructed, which were bootstrapped 1000 times and all major branches yielded values of 95–100%. These rodent orthologous relationships have been supported in my tree construction. Using the new rodent loci maps, we can see the synteny between mouse and rat *PSG* families is occurring before the major mouse *Psg* cluster (Fig:3.3.B).

The physical localisation of the 6 *PSG* genes in rat (*PSG38*, *PSG41*, *PSG36*,

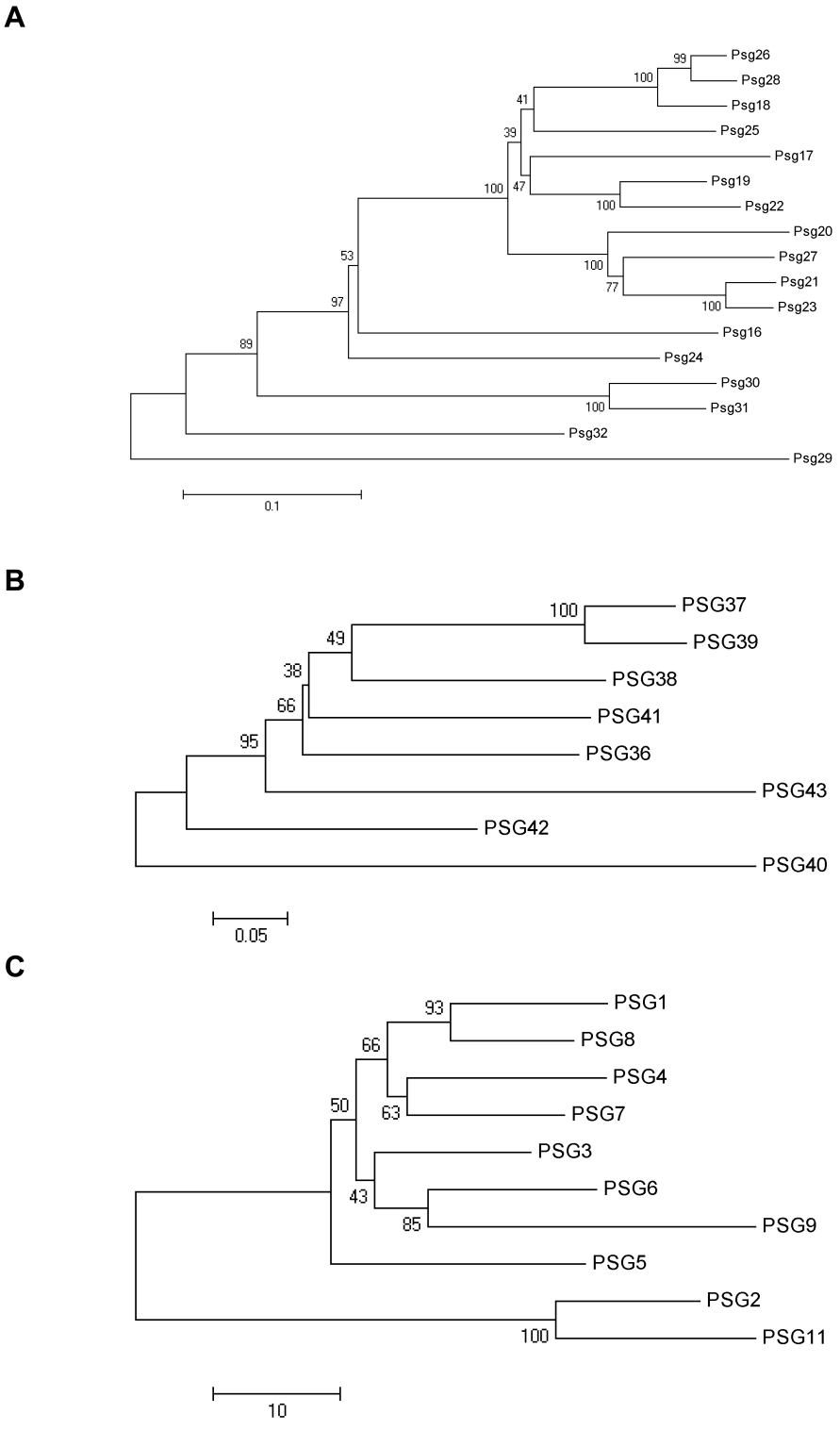


Figure 3.2: Phylogenetic trees of (A) murine CDS sequences, (B) rat CDS sequences and (C) human *PSG* AA sequences. Phylogenetic trees (Neighbour-joined pairwise comparison phylogenetic trees) were constructed using the MEGA4.0 software programme (http://www.megasoftware.net/). Data were bootstrapped 1000 times and all major branches yielded values of 95–100%. The scale bars represent 0.1, 0.5, or 10 nucleotide substitutions per site.

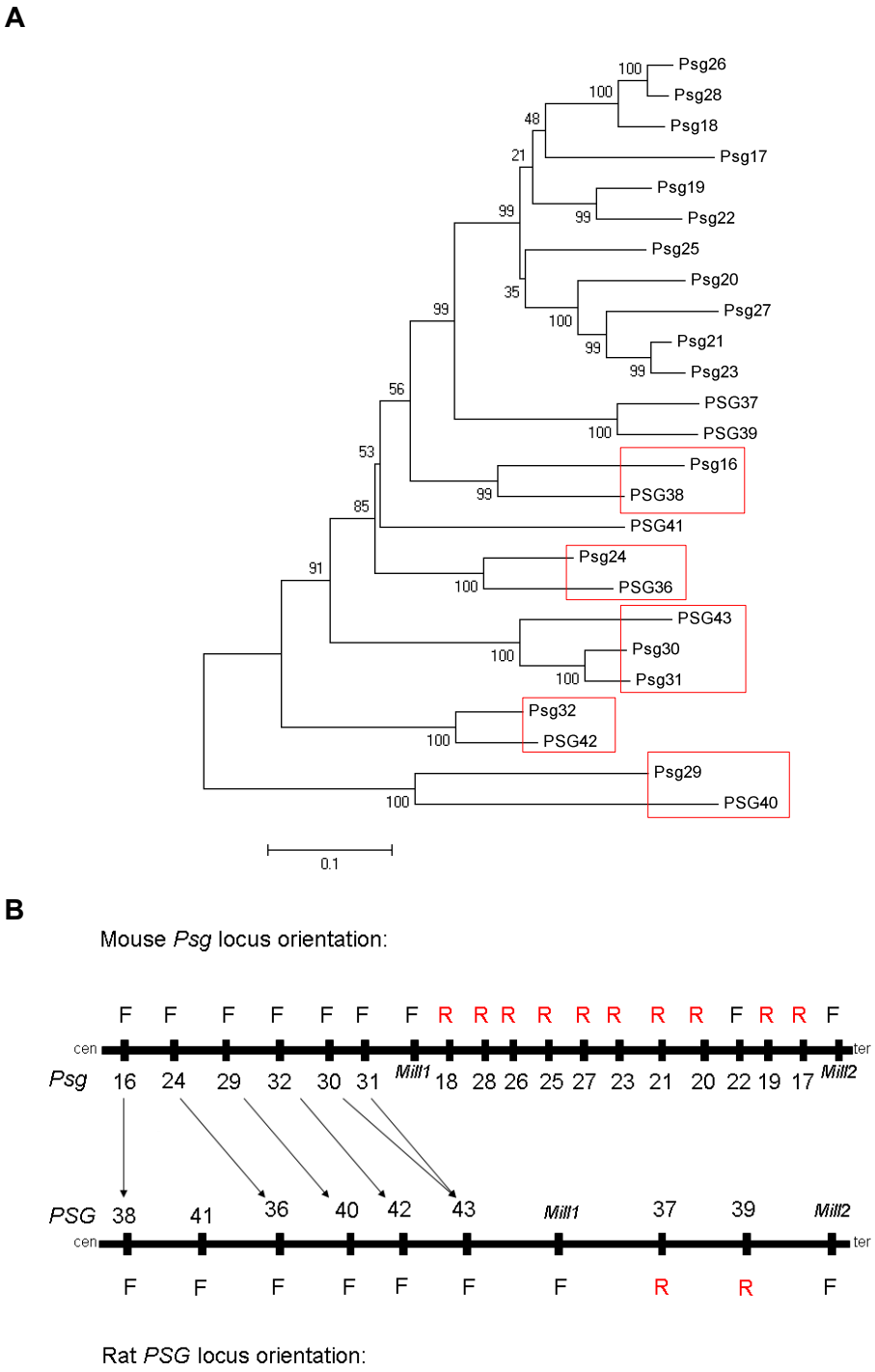


Figure 3.3: Rodent *PSG* orthologous relationships. (A) Phylogenetic tree of rodent *PSG* CDS sequences. Phylogenetic trees (Neighbour-joined pairwise comparison phylogenetic trees) were constructed using the MEGA4.0 software programme (http://www.megasoftware.net/). Data were bootstrapped 1000 times and all major branches yielded values of 95–100%. The scale bar represents 0.1 nucleotide substitutions per site. (B) Mouse and rat *PSG* loci synteny map showing orthologous relationships between these species before the *Mill1/2* flanked *PSG* cluster.

PSG40, *PSG42* and *PSG43*) and mouse (*Psg16*, *Psg24*, *Psg29*, *Psg32*, *Psg30*, and *Psg31*) showed a conserved order before the Mill1/2 cluster of *PSG* genes. Murine and rat genes localised within this chromosomal segment are shown (Fig:3.3.B). My phylogenetic analysis has revealed five orthologous relationships between rodent *PSGs*. These orthologous relationships can be seen as Mouse *Psg16* branches distinctly with rat *PSG38*. Mouse *Psg24* and rat *PSG36* cluster together, and there is also supporting evidence of this orthology in that both these *PSGs* contain 5 N domains. Mouse *Psg29* branches with rat *PSG40*, and mouse *Psg32* can be seen branching with rat *PSG42*. There is also orthologous relationships between mouse *Psg30* and *Psg31* with rat *PSG43*. All orthologous relationships occurs before the Mill1/2 cluster of *PSGs* in both rodent families and this orthology occurs in *PSGs* that are located on the forward strand. It is worth mentioning that rat *PSG37* and *PSG39* cluster on the same branch, and that this branch is closer to the murine *Mill1/2* flanked cluster of *Psgs* than to the rest of the rat *PSG* family members. There is a high confidence in the orthology demonstrated in the multi-species *PSG* phylogenetic tree as bootstrapping scores of 99-100% for each major branch shows that these branching points are robust. These updated locus and synteny maps will correct annotation in Ensembl database and facilitate future functional studies of this complex gene family.

3.1.3 Obtaining a *Psg* containing PAC clone - Mus musculus 129/Sv PAC library screen

To obtain a *Psg* containing PAC clone, a P1-derived Artificial Chromosome (PAC) library was screened for a *Psg23* positive clone. *Psg23* was used because it is located in the centre of the major murine *Psg* cluster, and is relatively close to *Psg22*. It was also chosen, as this major *Psg* cluster may be knocked out in the future. The RPCI-21 PAC Library has been constructed with female 129S6/SvEvTac mouse spleen genomic DNA (partially *MboI* digested) and was cloned between the *BamH1* sites of the pPAC4 vector [248]. The average insert size is 147 Kbp. The library consists of approximately 128,899 clones in 336 microtitre plates. The plate numbers run from 337 to 672. The PAC library has been gridded onto 22x22cm positively charged nylon

filters for hybridization screening purposes. Each filter contains 36,864 colonies which represents 18,432 independent clones spotted in duplicate in a 4x4 clone array. Seven filters cover the whole library. This provides a 6-9 fold coverage of the mouse genome. A 879 bp probe was designed approximately 2 kb upstream of *Psg23*, (Fig:3.4. A). Primers that amplified this region were designed and the PCR was performed using an 129/Sv genomic template. This PCR product was then cloned and sequenced. The probe was then excised from the positive clone using restriction endonucleases *Sal1* and *Nco1*. The RPCI-21 PAC Library was then screened using the probe hybridised to the 7 filters that cover the entire library. Positive signals were detected and analysed using the positive signal orientations to obtain correct clone numbers as per manufacturers instructions (Fig:3.4. C & D). 10 clones were picked based on signal strength. A mixture of weak, mid strength and strong positive signals was picked and ordered. The 10 clones were named PAC1 to PAC10 for ease of reference. The 10 positive PAC clones were cultured and prepped, using the Qiagen Large Construct kit as per protocol. *Psg23* specific primers were designed upstream and downstream of *Psg23*, these primers were used to determine which of the ten positive PAC clones contained *Psg23* sequence. The *Psg23* specific upstream primers amplified the correct product for *Psg23* in all ten PAC clones (Fig:3.4. E). Although only two of the PAC clones contained the positive *Psg23* sequence product for the *Psg23* downstream primers. PAC3 was chosen to continue with characterisation, as it contained *Psg23* sequences amplified by *Psg23* specific PCR primers. The PAC3 clone is clone 647-D4 in the RPCI-21 PAC Library. To determine which other *Psgs* were present on the PAC3, gene specific *Psg* primers were designed and PAC3 DNA was used as PCR template.

There are a number of other *Psg* family members located on PAC3, including *Psg25*, *Psg27*, *Psg23*, *Psg21*, *Psg20* and *Psg22* (Fig:3.4. F). To determine the exact *Psg* locus boundaries of PAC3, purified PAC3 DNA was sent to GATC (GATC Biotech, Germany) for End Sequencing [248] using the T7 and SP6 promoter sequencing primers on the pPAC4 plasmid backbone (Fig:3.4. B). End sequencing revealed that the region of the *Psg* locus present on PAC3 stretched from downstream of *Psg26*

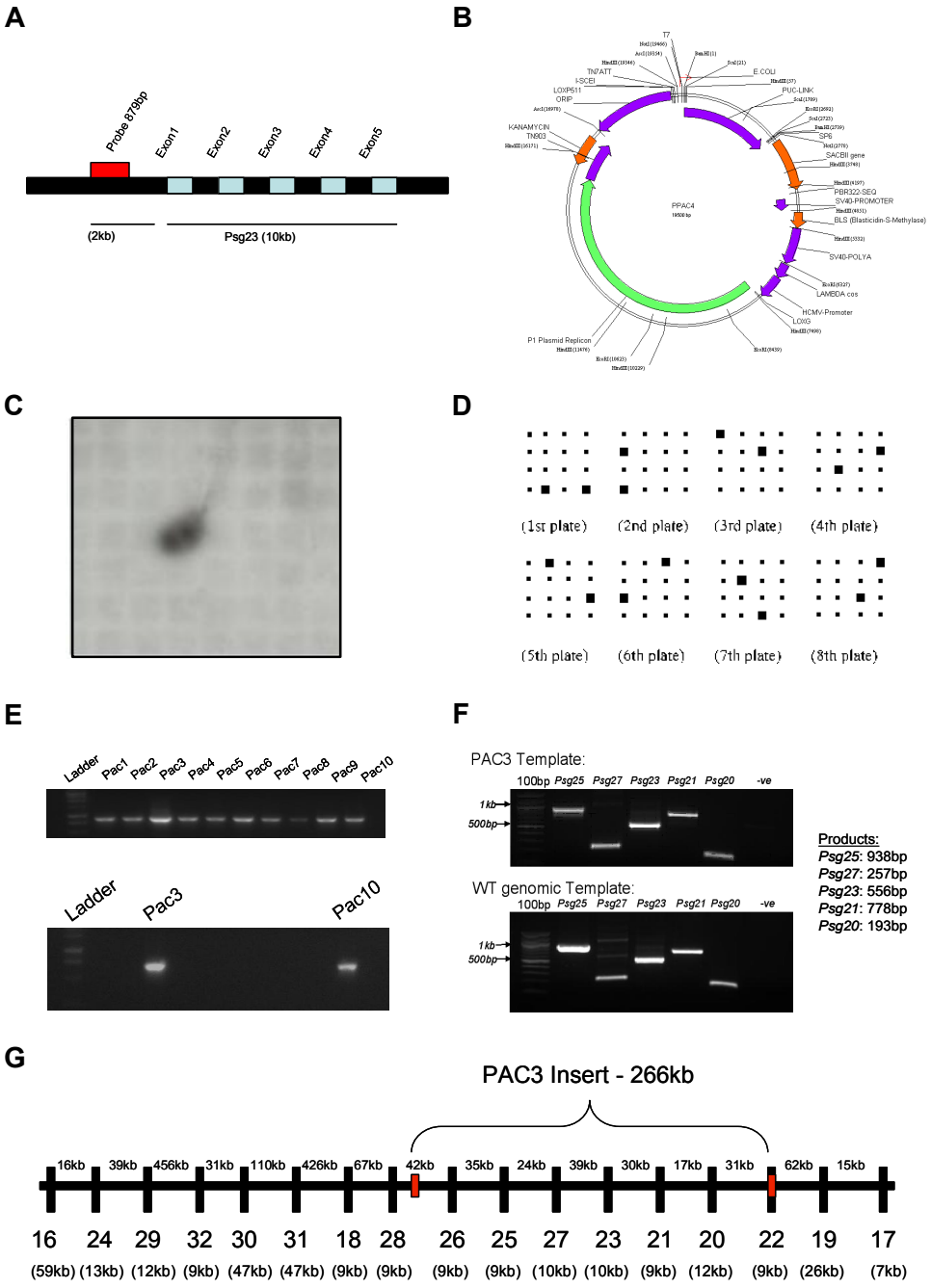


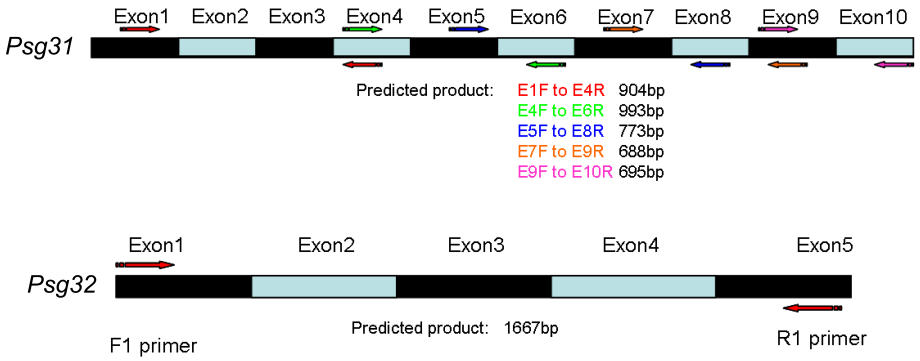
Figure 3.4: 129Sv PAC filter library screen. (A) *Psg23* specific probe located 2kb upstream of *Psg23*. (B) pPAC4 plasmid map. (C) Probe hybridisation positive signal. (D) PAC positive signal orientations. (E) *Psg23* specific primers upstream and downstream. (F) Psg specific primers for characterisation of PAC3 (G) PAC3 clone end sequencing.

to the middle of *Psg22* exon2 (Fig:3.4. G). The End sequencing of this PAC3 clone, has generated sequence data regarding the orientation of *Psg22* located on this clone. End sequencing has confirmed the inverted orientation of *Psg22* located on this PAC clone. This clone is derived from 129S6/SvEvTac mouse spleen genomic DNA, and the public genomic databases are based on C57BL/6J mouse genomic DNA, both of which have demonstrated that this *Psg22* inversion is not strain specific. More research needs to be undertaken to define whether this inversion event is common to all mouse strains. After characterising the PAC3 clone comprehensively, this clone was now ready to be used in the *Psg* Knockout vector construction as a source of isogenic homologous arms that will flank the KO vector to enable homologous recombination. Due to time constraints a *Psg* KO vector was not produced, although this PAC clone was used in other experiments during the course of this research.

3.1.4 Investigating the structure and expression of *Psg31* and *Psg32*

In McLellan et al, 2005 [148], two novel murine *Psg* genes were identified. Named *Psg31* and *Psg32* as per nomenclature [256], these transcripts were incorrectly annotated as a pseudogene (LOC381852/Gm5155) and a hypothetical gene (*Psg-ps1*), respectively on the NCBI databases. There was also conflicting data regarding exon number for *Psg31* in the public databases. *Psg-ps1* was previously considered to be a pseudogene, based on a point deletion at nucleotide position 30, downstream from the canonical *Psg* translational start site [106]. The open reading frame of *Psg32* initiates 105 bp upstream of the site of the mutation to an alternative ATG site. BLAST analysis of the public EST and Trace Archive EST databases yielded many mRNA clones that contain this region in addition to downstream exons. Hence, this gene is clearly expressed, and we now propose to rename *Psg-ps1* as *Psg32* hereafter. [148]. To provide a better understanding of these two genes, correct accession, sequence and expression data were generated to fully complete the *Psg* locus. Using the online BLAST alignment programme, all known *Psg31* sequences were BLASTed against a 2 Mb chromosome 7 sequence. This generated a full length genomic map of *Psg31* on the locus sequence and from these data I have been able to discern

A



B

Psg31:

Primer set 1: E1F-E4R:
TGGAAAGTGGTCCTTCCTTGGGAGACACTCAGATAAGAAG.....
.....GTGCAAATCTACAATGTCAC-GGCTGTAGAATCCTATATCCTCTTTG

Primer set 2: E4F-E6R:
GCCTACCACTGCCCAAATAACCATTGAATTAGTGCCACCC.....
....AAAATTGAATCACTTCCTCA-CACTTTTCCCTACAGCAACTTTTT

Primer set 3: E5F-E8R:
AGCTGTACATGGACACATCCCTTTCTACTTGCTACCAACTT.....
....ATATACACCAATGGATCCAT-CAGTGACATCAATCATCGCC

Primer set 4: E7F-E9R:
TGTTGGAGGGGAAAGTGTTCTTCTACTGGTTCACAATCTC.....
..... TCCAAACTCCAAGGACTCTG-ACTGAACCTTGCAGTGGAGC

Primer set 5: E9F-E10R:
CAAAGGTGTGATTGCAGAGGAGAAATCTGAGCTCATCCAA.....
..... GCTGACCCTGGAGTCTCCAT-GCTTATTGAAGAGCCAACGG

Psg32:

Primer set 1:
GGAAGTGTTCCTGCCTGAGAGAACACTCAGATCAGAAGAA.....
.....GAAAATTGTCAGTTACCACT-CAGGTACAGCCACCATTGTG

Figure 3.5: Expression of *Psg31* and *Psg32*. (A) Graphical representation of primer sites in *Psg31* and *Psg32* and predicted gene structure. (B) Primer sequences (red) and intial sequencing returned from cloned and sequenced RT-PCR products. *Psg31* and *Psg32* are expressed in E15 CD1 placenta tissues.

that there are 10 exons contained in *Psg31*. The current model of *Psg31* evolution is that *Psg31* evolved from a duplication of the whole of the ancestral *Psg30* gene followed by a subsequent internal duplication of the N1 domain [146]. Predicted domain structures of *Psg31* and *Psg32* are shown (Fig:1.6.A). There was also no data regarding whether these novel *Psg* transcripts were expressed in murine placenta. I wanted to ascertain the correct exon number in *Psg31*, and to discern whether *Psg31* and *Psg32* are expressed in murine placenta. Gene specific primers were designed to amplify overlapping sequences in *Psg31* and a specific primer set to amplify the whole *Psg32* CDS. (Fig:3.5.B). E15 placental cDNA was synthesised, and RT-PCR was performed. Cloning and sequencing of RT-PCR products confirmed that *Psg31* has ten exons and that *Psg31* and *Psg32* are expressed in E15 placental cDNA. (Fig:3.5.A& B) shows primer locations on these genes and the positive sequences. I found that the previously uncharacterised mouse *Psg31* and *Psg32* genes were expressed in E15.5 murine placenta. This expression data is important as it shows that there are 17 transcribed *Psg* genes in the mouse. It also gives us a better understanding of the structure of both *Psg31* and *Psg32*. *Psg32* is structured like the majority of the murine *Psgs*, containing five exons contributing to three Ig-V-like domains and a Ig-C-like domain. *Psg31* is the largest of the murine *Psgs*, containing 10 exons which contribute to 8 Ig-V-like domains and a Ig-C-like domain. This *Psg31* gene, which is closely related to *Psg30* but, uniquely amongst murine *Psg* genes, has a duplicated N1 domain.

3.1.5 Differentiated TSC as a model for endogenous *Psg22* expression

It has been reported previously that the primary site of murine *Psg* expression occurs in TGC [148, 49, 165], although at present there is no immortal murine TGC line and there is a distinct lack of trophoblast cell lines that fully recapitulate the behaviour of early placental trophoblast [55]. To obtain a source of endogenous *Psg* expression *in vitro*, a cell line model expressing endogenous *Psg* needed to be established. As mentioned previously, TSC will differentiate primarily into TGC and SpT [44]. To determine whether *Psg22* expression is dynamically regulated during trophoblast

differentiation, we used TS-EXE and TS-GFP trophoblast stem cells as a model system. I obtained two trophoblast cell lines, TS-EXE and TS-GFP as described elsewhere [21, 44] and attempted to differentiate these TSC lines into predominantly TGCs.

It has been reported that retinoic acid (RA) contributes to TGC differentiation with the suppression of the SpT formation. TSC cells treated with RA for 48 hours exhibited attenuated growth and extensive morphological change [101]. It has also been reported that RA, specifically 9-cis retinoic acid upregulates *PSG5* expression in humans through a functional Retinoic Acid Responsive Element (RARE) motif shared by all human *PSG* genes [200]. So using RA as a tool to differentiate TSC, TSC cells were treated with 5 µM retinoic acid (both all-trans retinoic acid (ATRA) and 9-cis retinoic acid (9cisRA) for 24 hours and 48 hours respectively. EtOH was used as a vehicle control. TSC cells were also subjected to a conditioned media withdrawal (withdrawal of FGF4, heparin and MEF conditioned medium) protocol of differentiation as described [21]. RNA was extracted and cDNA synthesised as described in materials and methods. RT-PCR was performed using this cDNA as template to examine the molecular markers of TGC differentiation. Marker genes used to determine differentiation were: *Eomes*, a TS cell marker, *TpbpA*, a SpT marker, and *Prolactin2 (Pl2)*, which is a TGC specific marker gene. Primers used are described in [44], and are listed (Table 2.3.).

The trophoblast marker *Eomes*, has a low level of expression in both TS cell lines treated with RA 24 and 48 hours and *Eomes* is highly expressed in both TS cell lines treated with vehicle control after 48 hours (Fig:3.6.A). Even though *Eomes* is still being expressed in the RA treated cells, a proportion of the cell population has been differentiated into TGCs as can be seen from the expression of *Pl-2*, 24 and 48 hours post treatment. As expected there is some expression of the SpT marker *TpbpA*, as these cell lines do not produce a pure population of TGC, as can be seen with the low level expression of SpT marker *TpbpA* after 24 and 48 hours respectively. The vehicle control treated TSC shows that these TSC are differentiating towards the SpT fate as there is faint expression of *Pl2* but high expression of *TpbpA* post treatment. From this experiment we can see that the RA treated TSC are indeed differentiating towards a

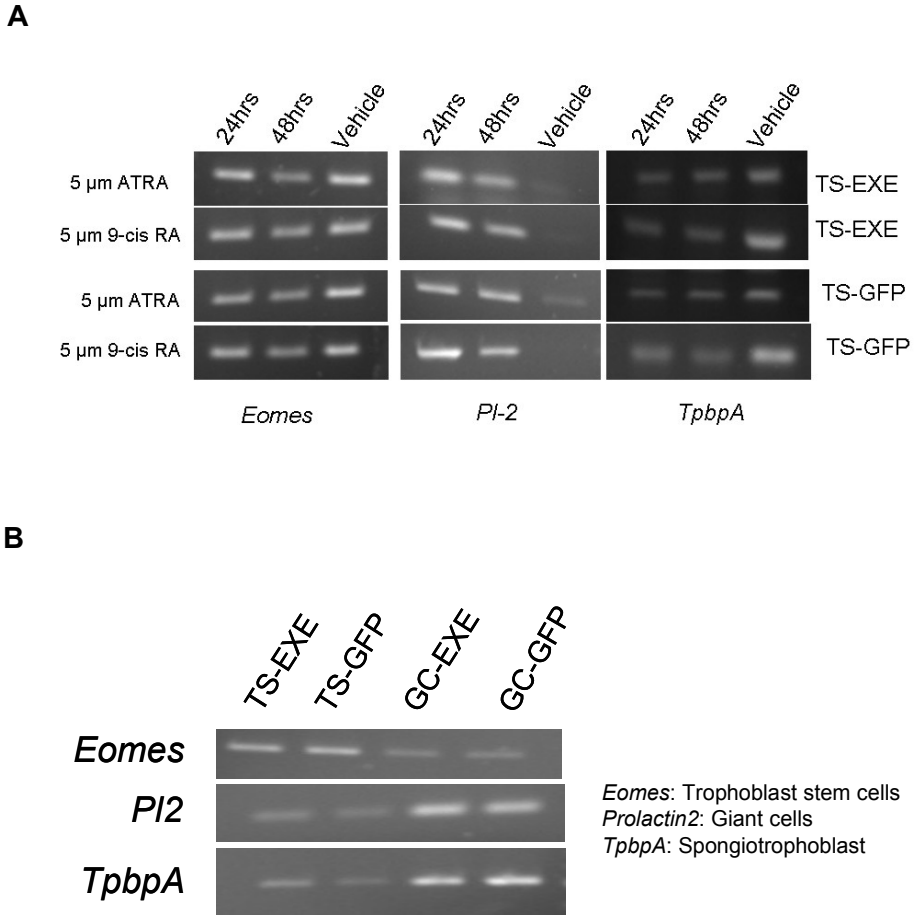


Figure 3.6: Molecular characterisation of differentiated TSC. (A) RT-PCR of TSC differentiation markers expressed in retinoic acid treated cells. Differentiation markers used were: *Eomes* - trophoblast marker, *Pl-2* - TGC marker and *TpbpA* - SpT marker. (B) RT-PCR of undifferentiated TSC (TS-EXE, TS-GFP) and 6 day conditioned medium withdrawal differentiated TSCs (GC-EXE, GC-GFP) using differentiation marker primers.

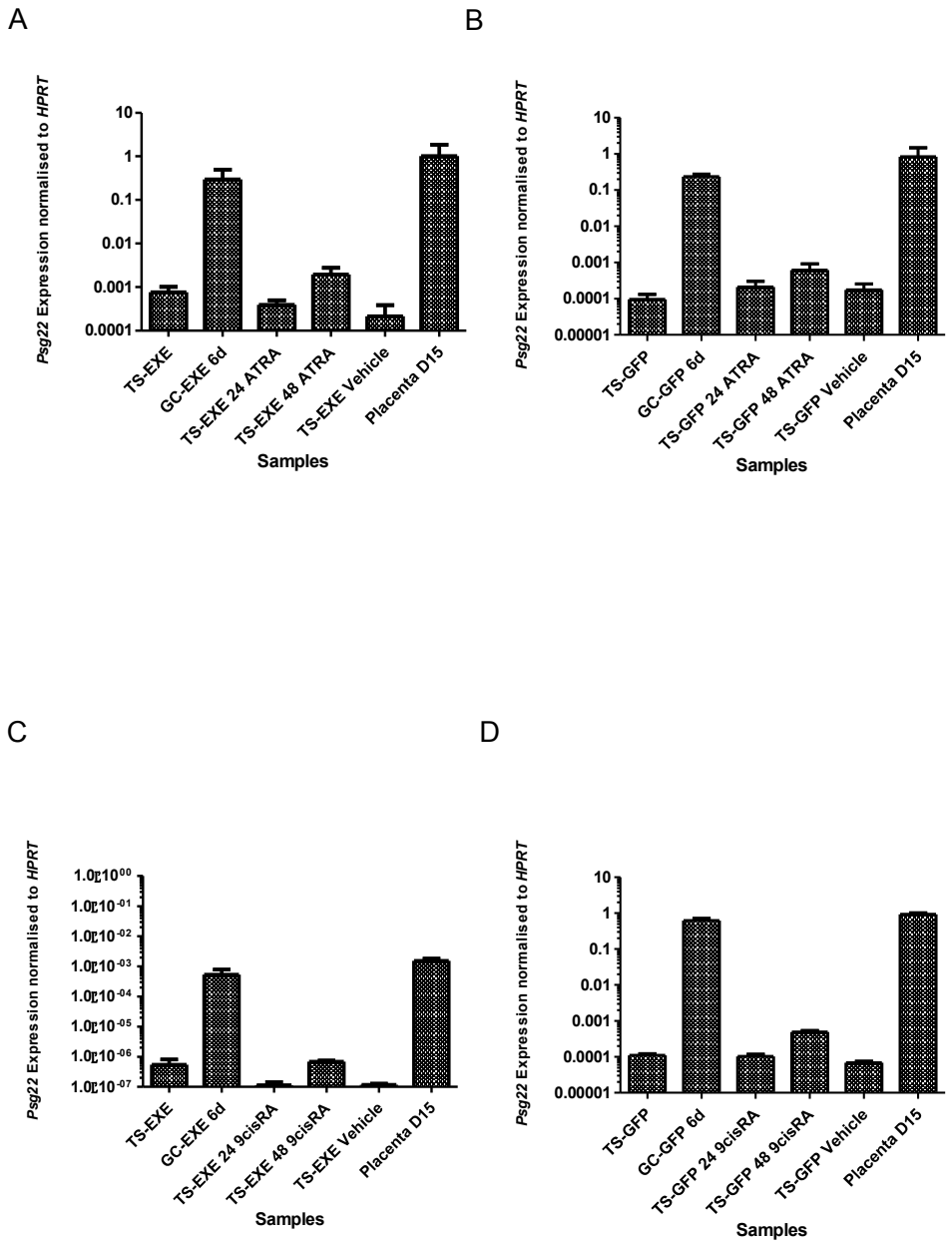


Figure 3.7: Differentiated TSC model of endogenous *Psg22* expression. Relative quantification of *Psg22* expression normalised to *Hprt* expression in TS-EXE (A&C) and TS-GFP (B&D) cell lines. *Psg22* expression is induced by 6 day FCM media withdrawal and retinoic acid (5 µM all trans retinoic acid (A&B) and 9-cis retinoic acid (C&D) for 24 and 48 hour treatments). (n=3).

TGC fate, although there is some proportion of the population differentiating towards a SpT cell fate. The 6 day conditioned medium withdrawal protocol has differentiated TSC to TGC and SpT cells as can be seen in the relatively high expression of *Pl2* and *TpbpA* in the cells which have undergone the 6 day conditioned medium withdrawal protocol (Fig:3.6.B). Both the RA treatments and the 6 day conditioned medium withdrawal protocol has produced a TGC population *in vitro*.

The relative levels of *Psg* expression in both these protocols needed to be assessed to determine which protocol produced a similar level of *Psg* transcription to dissected placental tissues. To assess *Psg* expression levels, qRT-PCR was used to compare *Psg* expression in RA treated TSC, 6 day conditioned medium withdrawal TSC, and E15 CD1 placenta. *Psg22* qRT-PCR primers (Table 2.7.) were used. The relative *Psg22* expression in RA treated TSC, (ATRA or 9cisRA), 24 and 48 hours post treatment is shown (Fig:3.7.A-D). Low levels of *Psg22* expression was detected in undifferentiated TSC populations. This is consistent with the observation that a small percentage of TSC undergo differentiation to TGC even in the presence of FGF4 [21].

The RA treatment has induced *Psg22* expression but only marginally compared to undifferentiated TSC, even after 48 hours post-treatment. Interestingly, there was not much difference in the *Psg22* induction levels shown between the 9cisRA and the ATRA, which is surprising given that [257] stated that *Trans*-activation analyses show that although all three RXR receptors respond to a variety of endogenous retinoids, 9-cis RA is their most potent ligand and is up to 40-fold more active than ATRA. *Psg22* expression in these RA treated cell lines has failed to induce *Psg22* to endogenous placental levels. In contrast, the 6 day conditioned medium withdrawal differentiation protocol induced *Psg22* expression to levels that are comparable to endogenous *Psg22* expression in E15 placental tissues. From this expression data, it is evident that RA does induce *Psg22* expression marginally after 48 hours. Higher doses and longer post-treatment time points may boost RA induction of *Psg22* expression. The 6 day conditioned medium withdrawal TSC differentiation protocol demonstrated the best ability to mimic placental endogenous

Psg22 expression in TGC cell lines, and is the differentiating protocol used in the rest of this work.

3.1.6 Expression survey of PSG transcript abundance - cloning screens

Comparative expression studies of multigene families provides important insights into biological processes that have potential or known importance for our understanding of the mechanisms of development. I undertook a PSG expression survey of a variety of trophoblast and TGC derived tissues. Previous studies have shown that *Psg21* and *Psg23* gene transcripts together constitute the bulk of *Psg* gene expression in the SpT [191, 148]. It has also been shown that *Psg22* is the most abundant transcript in TGC [49]. To determine whether specific *Psg* gene transcripts similarly dominate in TGC derived from differentiated TSC lines, two degenerate primer sets were designed to amplify all known mouse *Psgs* [49], (Table 2.2.). As previously described, RT-PCR amplicons were cloned into pSTBlue1 cloning vector and positive clones were sequenced. I investigated relative *Psg* transcript frequency in two TSC lines - TS-EXE and TS-GFP. cDNA was synthesised from extracted total RNA, and used as template in RT-PCR reactions. 10 positive clones from each primer set amplifying template from each TSC line was sequenced and it was found that in the TS-EXE cell line, *Psg22* was the most abundant transcript, although there was also a variety of other *Psg* transcripts expressed in this TSC line, including *Psg16, Psg20, Psg23, Psg27* and *Psg28* (Fig:3.8.A). In TS-GFP cell line, the most abundant *Psg* transcript was *Psg27*, although, as for TS-EXE cell line, there was also a variety of *Psg* transcripts expressed, including *Psg17, Psg20, Psg22/25, Psg27,* and *Psg28*. In contrast, TGC (GC-EXE and GC-GFP) that have been differentiated from these TSC lines, clearly show that *Psg22* is the most abundant *Psg* transcript in differentiated TGC (Fig:3.9.A&B). 80% of clones sequenced in both GC cell lines were either *Psg22* or *Psg25* transcripts. This *Psg* expression survey of two TSC lines and their derived TGCs has shown that there is a variety of *Psg* transcripts expressed in undifferentiated TSC but *Psg22* is the most abundant transcript present in differentiated TGC. This data is consistent with previous results demonstrating predominant *Psg22* expression in

dissected TGC [49].

To compliment the results of our survey of *Psg* expression in undifferentiated TS cells and their differentiated TGC, I employed the same experimental procedures in isolated mouse C57BL/6J E5 blastocysts, and E11 blastocyst outgrowths. As described in the introduction, E5 blastocysts contain predominately TE, including an abundance of TSC, whereas by E11, the blastocyst outgrowth is predominantly comprised of differentiated TGC. Similar to undifferentiated TSC lines, E5 blastocysts contain a variety of *Psg* transcripts, the majority being *Psg22/Psg25*, but also including *Psg16, Psg17, Psg18, Psg20, Psg23, Psg27*, and *Psg28* (Fig:3.10.A). A *Ceacam9* transcript was also amplified with the PSGFR primer set although this is unsurprising as CEACAMs exhibit homology to PSG sequences in a variety of species and this primer set consists of degenerate sequences. In contrast to E5 blastocysts, the major *Psg* transcript in E11 blastocyst outgrowths is *Psg22* as in differentiated TGC, with 65% of clones sequenced being *Psg22*.

3.1.7 Investigating *Psg* expression in the gasterointestinal tract

Early research on tissue-specific expression of *Psg*, indicates that murine *Psg* expression is detected exclusively in TGC and SpT of the placenta [129, 130, 131]. Although expression of *Psg18* was described [192], in follicle-associated epithelium (FAE) above Peyers' patches (PP) in the GIT, and more recently, the report of a brain specific transcript of *Psg16* by [193] led to the hypothesis that *PSGs* were not expressed exclusively in the placenta [129]. To determine whether murine *Psg* expression was located elsewhere in the GIT, a *Psg* expression survey of eight different GIT tissue samples was undertaken. The same primer sets to amplify all known murine *Psg* as utilised in the TSC/TGC and blastocyst expression surveys were employed to assess *Psg* expression in both male and female GIT. Eight tissues were used in this survey covering the length of the GIT, including oral cavity, esophagus, stomach (pylorus), small intestine, ileum, caecum, and rectum. Tissues were dissected from two male and two female CD1 mice, RNA was extracted and cDNA was synthesised. RT-PCR amplicons were cloned into pSTBlue1 cloning vector and positive clones were

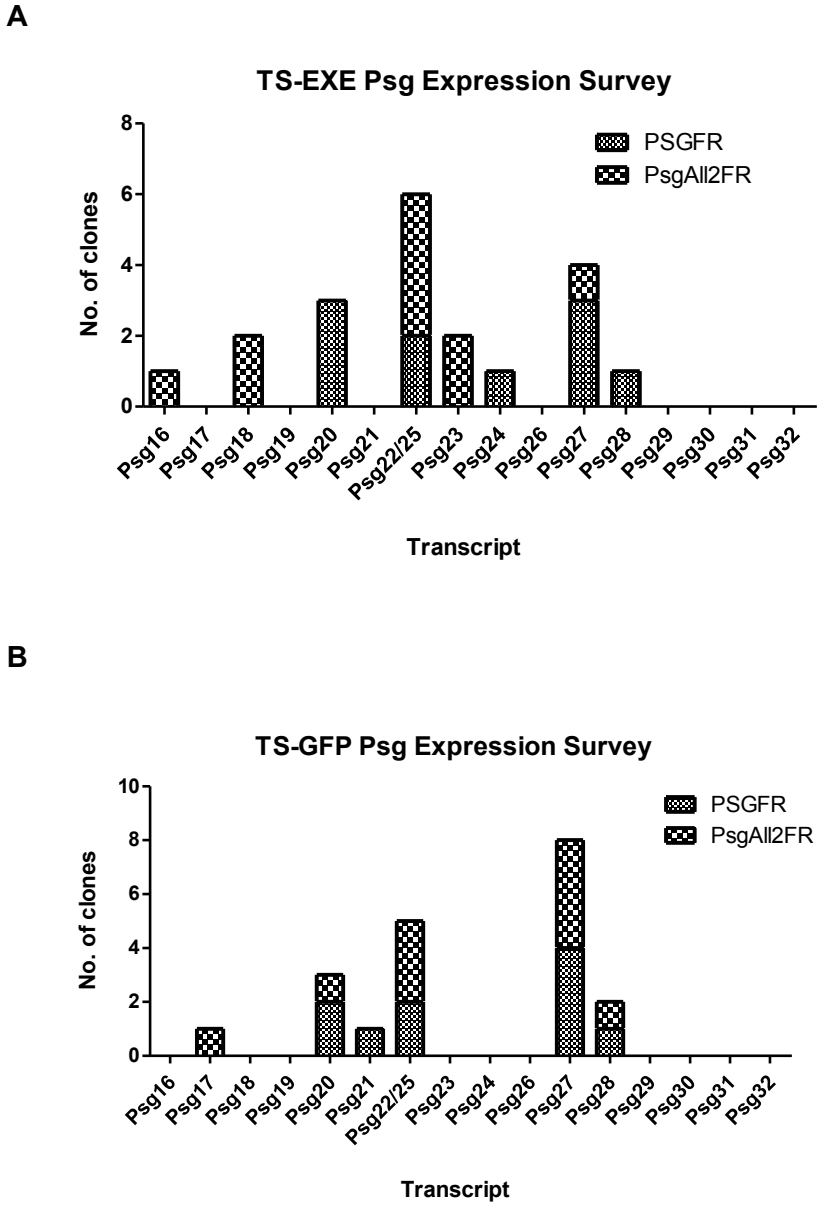


Figure 3.8: Murine *Psg* expression survey of two TSC lines (A) TS-EXE and (B) TS-GFP. RT-PCR performed with TSC cDNA, using primer sets PSGFR and PsgAll2FR that amplify all known murine *Psg*. 20 clones for each cell line were sequenced. Returned sequences were BLASTed against predicted *Psg* amplicons.

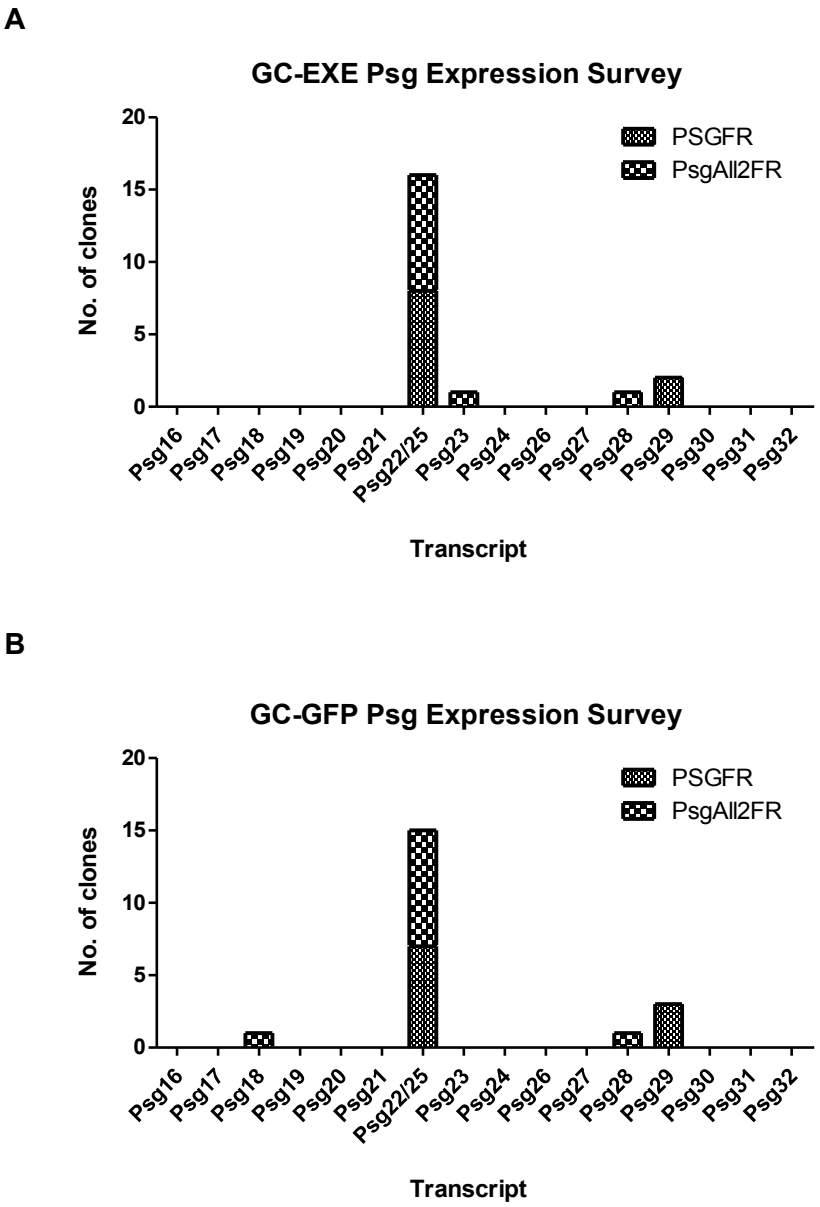


Figure 3.9: Murine *Psg* expression survey of two TGC lines (A) GC-EXE and (B) GC-GFP. RT-PCR performed with TGC cDNA using primer sets PSGFR and PsgAll2FR that amplify all known murine *Psg*. 20 clones for each cell line were sequenced. Returned sequences were BLASTed against predicted *Psg* amplicons.

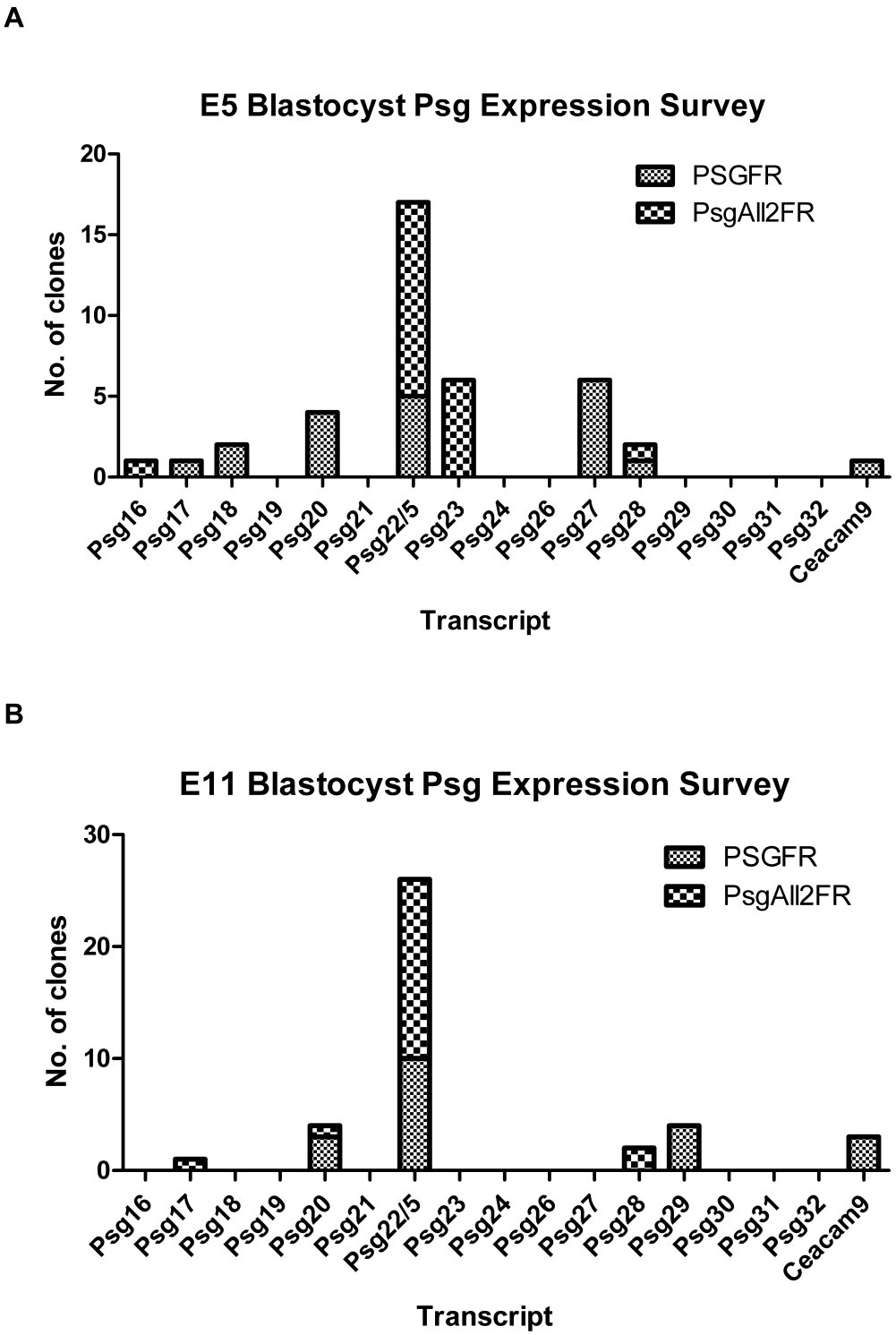


Figure 3.10: Murine *Psg* expression survey of (A) E5 blastocysts and (B) E11 blastocyst outgrowths. RT-PCR performed with blastocyst cDNA using primer sets PSGFR and PsgAll2FR that amplify all known murine *Psg*. 20 clones for each primer set amplifying each blastocyst stage were sequenced. Returned sequences were BLASTed against predicted *Psg* amplicons.

sequenced. Four positive clones from each tissue sample, for each primer set were sent to GATC (GATC Biotech, Germany). Returned sequences were BLASTed against predicted *Psg* amplicons to determine which transcript was present in each clone using the online BLAST software. The *Psg* expression survey has shown that a variety of *Psg* transcripts are expressed in all GIT tissues investigated (Fig:3.11.A& B). Both primer sets were able to amplify a variety of *Psg* transcripts. The PsgAll2FR primer set amplified seven of the seventeen mouse *Psg*, including *Psg18*, *Psg21*, *Psg22*, *Psg23*, *Psg25*, *Psg26* and *Psg28*. The PSGFR primer set amplified fourteen of the seventeen *Psgs*. The *Psg* amplified by this primer set were *Psg16*, *Psg18*, *Psg19*, *Psg20*, *Psg21*, *Psg22/25*, *Psg24*, *Psg26*, *Psg29*, *Psg28*, *Psg29* and *Psg31*. *Psg21* was found to be the most abundant transcript amplified by the PsgAll2FR primer set in all tissues, comprising of 40% of clones sequenced. *Psg22/25* transcripts were the major transcripts amplified by the PSGFR primer set. Moreover, *Psg31* was also demonstrated to be expressed in rectal tissue, supporting our results that *Psg31* is expressed and is a functional member of the *Psg* family (Fig:3.11.B). Synder *et al*, (2001), states that PSG released by the placenta plays a pivotal role in the induction of the Th2 response [160]. Kawano *et al* (2007), suggests that this hypothesis could apply to the mucosal immune system as well [192]. The bias toward Th2 response in PPs is essential for the production of secretory IgA and the tolerogenic response to commensal bacteria as well as food antigens [258]. The suggestion that *Psg* expression throughout the GIT is involved in the promotion of oral tolerance, complements their role as immunomodulators in the placenta.

To further elucidate the expression of *Psg* in the GIT, qRT-PCR was utilised to quantify the relative levels of *Psg* expressed in the GIT in relation to placental *Psg* expression levels. qRT-PCR was performed, using the degenerative PsgAll2FR primer set. Esophageal, ascending colon and E15 placental tissue was used as template. Results were described as mean *Psg* expression relative to mean *Hprt* expression. Normalisation of expression levels to the housekeeping gene, (*Hprt*), was used to avoid discrepancies caused by variations in input RNA or in reverse transcription efficiencies. The results show that *Psg* is expressed in the GIT and can be quantified

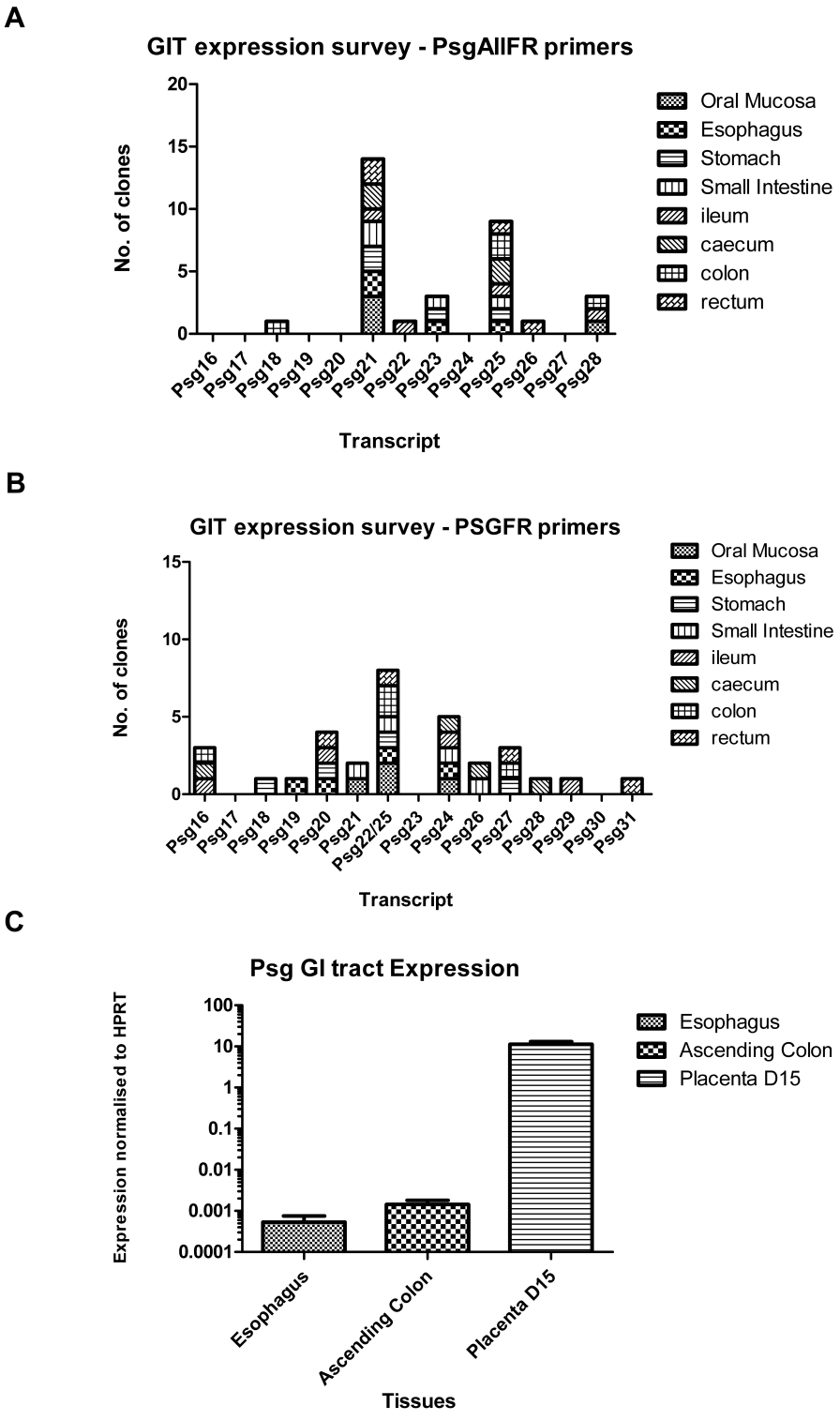


Figure 3.11: Murine *Psg* expression survey of gastrointestinal tract (male and female CD1 mice) using primer sets (A) PsgAll2FR and (B) PSGFR, that amplify all known murine *Psg*. Four clones were sent from each tissue sample for each primer set to be sequenced. (C) Relative quantitative expression of total *Psg* in murine esophageal and ascending colon tissue samples in contrast to E15 placental expression using PsgAll2FR primer set. (n=3).

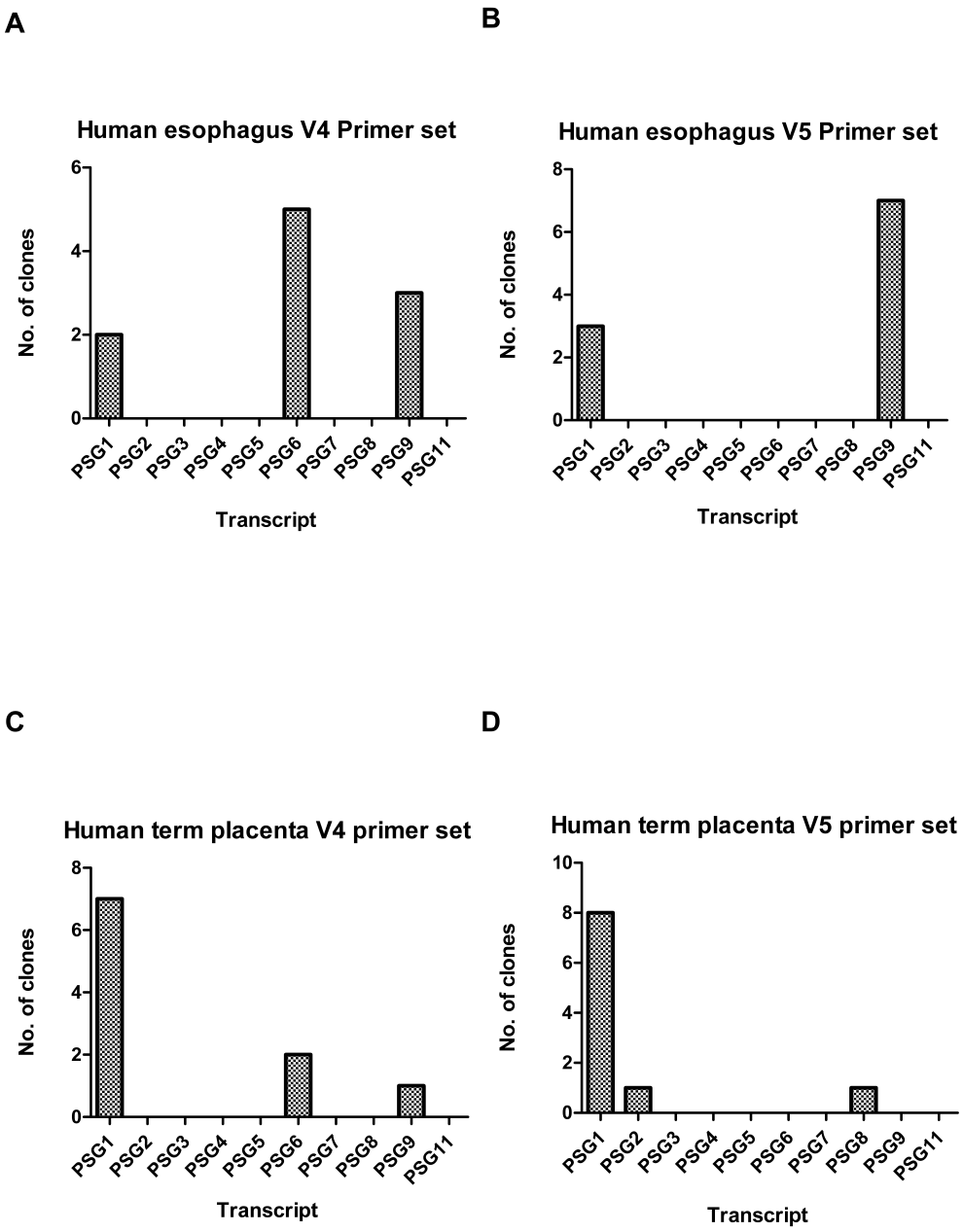


Figure 3.12: Human *PSG* expression survey: (A&B) esophagus and (C&D) term placenta. RT-PCR performed with esophagus and term placental cDNA and two primer sets V4 and V5 amplify all human *PSG* transcripts. 10 clones from each primer set sequenced.

reliably. The relative levels of *Psg* expression in the esophagus and ascending colon were considerably lower, about 4 orders of magnitude, than *Psg* levels found in the placenta (Fig:3.11.C).

To see if these results could be reproduced in human tissue samples, a human *PSG* expression survey was undertaken to analyse *PSG* expression in the Human GIT. Two degenerate primer sets, Human PSG V4 and Human PSG V5, (Table 2.2), were designed to amplify all *PSG* sequences. Human esophagus and human term placental cDNA were used as templates. As per previous expression surveys, the RT-PCR amplicons were blunt cloned into pSTBlue1 cloning vectors. From each tissue, 10 positive clones were sequenced for each primer set, and returned sequences were compared against predicted *PSG* amplicons to assess which *PSG* transcript was present in each clone. *PSG* expression was detected in the human esophagus. *PSG1, PSG6* and *PSG9* transcripts were amplified by both primer sets in esophageal tissue samples, with *PSG9* being the most abundant *PSG* transcript detected (Fig:3.12.A& B). In comparison, five out of 10 *PSGs*: *PSG1, PSG2, PSG6, PSG8* and *PSG9* were found to be expressed in term placenta using the same primer sets (Fig:3.12.C& D). *PSG1* was found to be the most abundant transcript found in these placental samples. Whether the human *PSGs* have the same levels of expression in the esophagus as found in the placenta needs to be investigated.

3.1.8 Quantitative expression of *Psg* in trophoblastic lineages

McLellan *et al* states that because all mouse *Psg* genes originated from a common ancestor, and through duplication and subsequent divergence expanded into a multigene family, the investigation as to whether the expression patterns have also diversified is relevant to determining the selective forces underlying *Psg* gene family expansion [146]. Initial investigations of *Psg* expression was performed in TSC lines and their differentiated TSCs. As seen in the TSC and TGC expression surveys, *Psg22* is upregulated when TSC cells are differentiated towards a TGC fate. I wanted to confirm this differential upregulation of *Psg22* and other highly transcribed murine *Psg* using relative qRT-PCR. *Psg* gene specific primers were designed using Primer-

BLAST software to ensure primer specificity, and TSC and their differentiated TGC cDNA were used as templates. Three biological replicates of each cell line were evaluated, using three technical qRT-PCR replicates. Normalisation of expression level to the housekeeping gene, *Hprt*, was used to avoid discrepancies caused by variations in input RNA or in reverse transcription efficiencies. Dissociation curves for the PCR products demonstrated a single specific peak indicating absence of non-specific amplification.

In this study, the expression of four murine *Psg* genes; *Psg19*, *Psg21*, *Psg22* and *Psg23* was quantified. These *Psg* family members were chosen, as *Psg22* has the highest levels of *Psg* expression in the first half of pregnancy, while *Psg21* and *Psg23* share the highest expression levels in the second half of pregnancy [49, 191]. *Psg19* was also chosen as it is the closest *Psg* family member to *Psg22* as can be seen in their phylogenetic branching (Fig:3.2.A). All four *Psg* genes quantified show similar increased *Psg* expression patterns when TGCs are differentiated from TSC (Fig:3.13.A-D). For each *Psg* quantified there is an increase of *Psg* expression in TGCs compared to their TSC derivatives. *Psg19*, *Psg21*, and *Psg23* all show approximately a 4 fold increase of expression upon differentiation (Fig:3.13.A, B,&D). These three *Psgs* show the lowest levels of increased expression. In contrast, the greatest increase of *Psg* expression upon differentiation is *Psg22*, (Fig:3.13.C), where there is a 6 fold increase in *Psg22* expression, the greatest increase seen in the GC-GFP cell line. This data reinforces the finding of the induction of *Psg22* expression in TGCs upon differentiation, as shown in the above *Psg* expression cloning screens of undifferentiated and differentiated TSC (Figs:3.8. & 3.9.). The fact that *Psg22* has the highest levels of expression in differentiated TGCs in comparison to the other three *Psg* quantified supports the hypothesis of a specific function for *Psg22* in TGCs in the early stages of placental development.

Following on from the quantification of *Psg* expression in TSC and TGCs, I investigated *Psg* expression patterns in a variety of trophoblastic lineages. TGCs and EPC tissue samples were dissected from E10 and E11 CD1 mice as described elsewhere [27]. Full placental tissue samples, where only the SpT compartment supports *Psg*

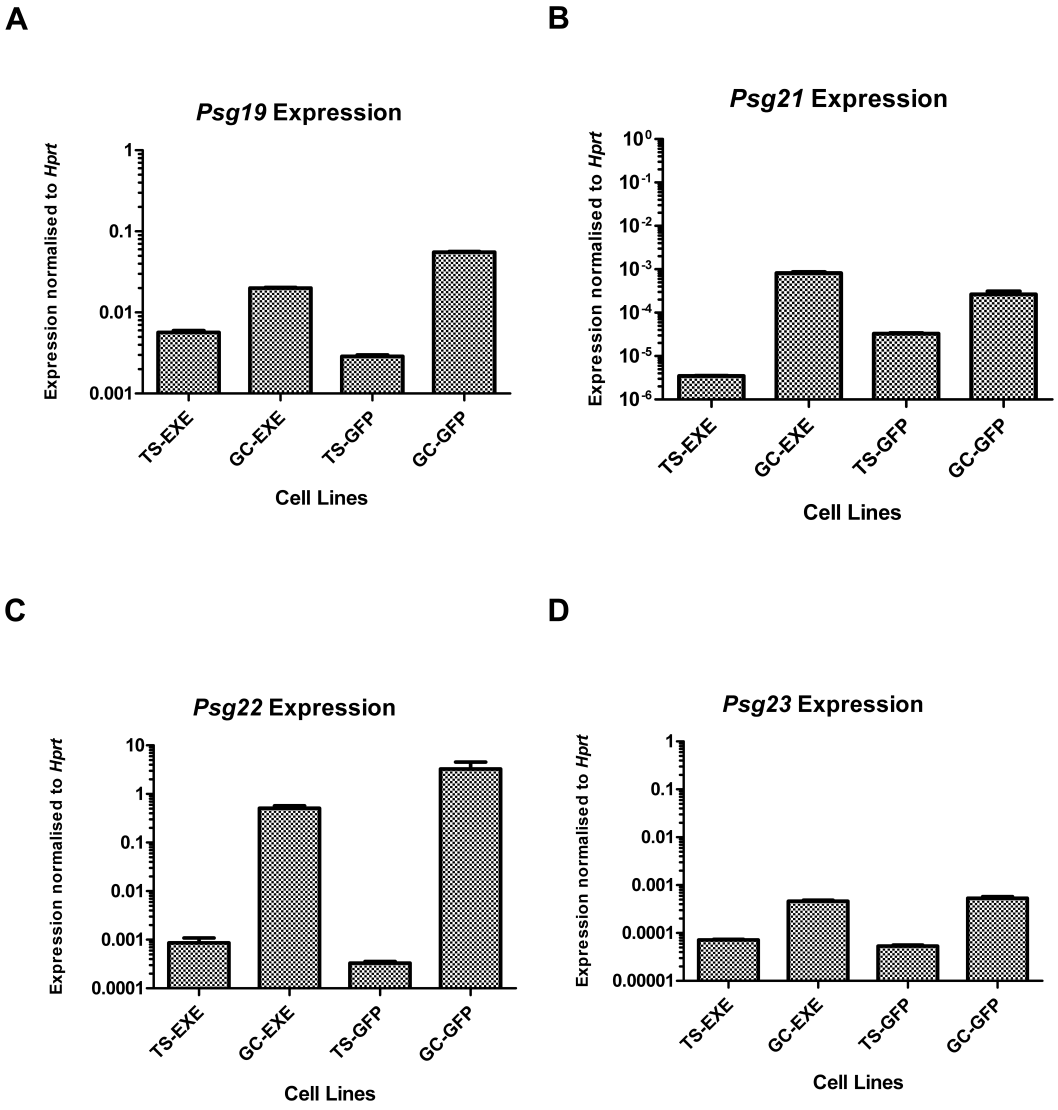


Figure 3.13: Relative quantification of *Psg* expression in TSC lines (TS-EXE and TS-GFP) and 6 day differentiated TGCs. *Psg* expression is normalised to *Hprt* expression. This demonstrates the induction of *Psg* expression when TSC are differentiated to TGCs. (A) *Psg19* expression, (B) *Psg21* expression, (C) *Psg22* expression and (D) *Psg23* expression. (n=3).

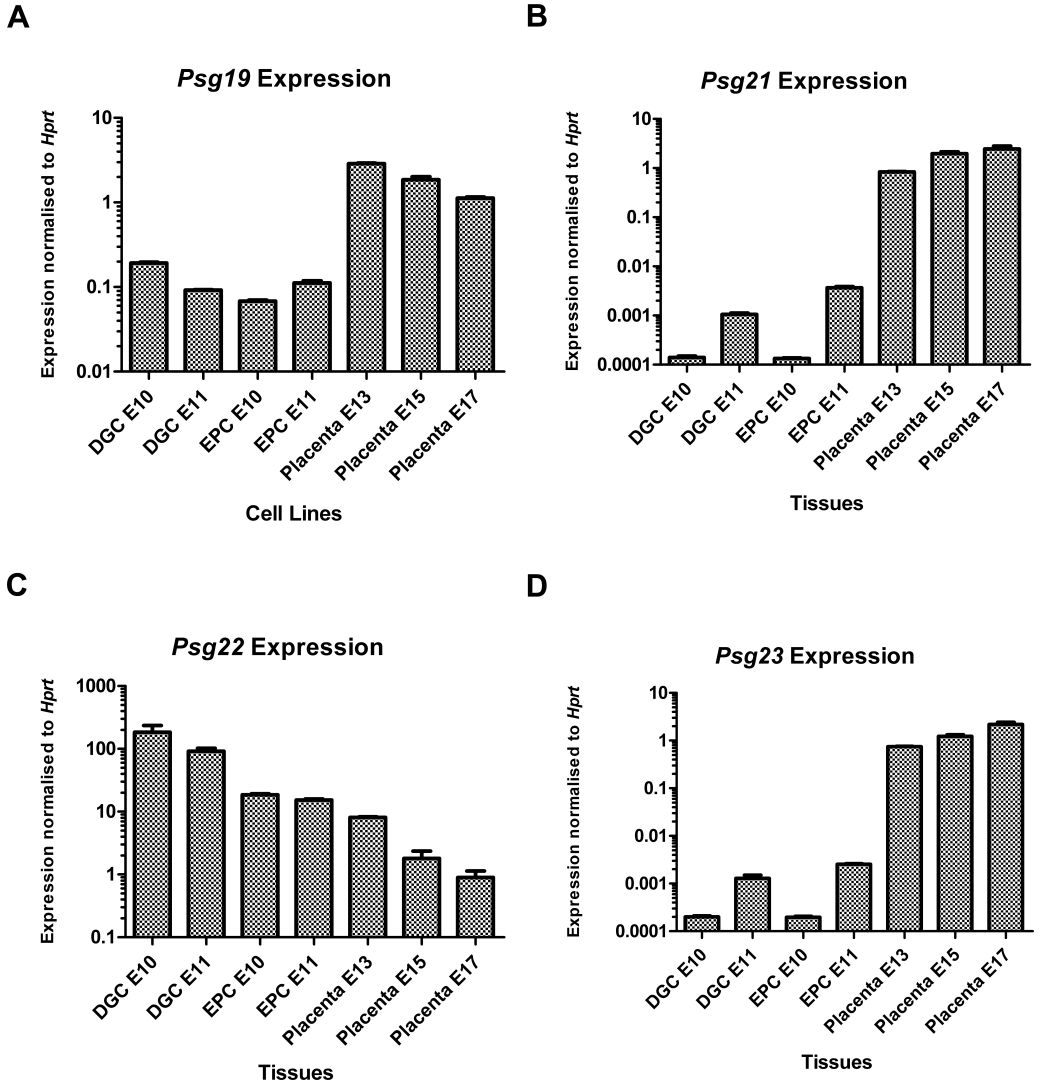


Figure 3.14: Relative quantification of *Psg* expression in trophoblastic lineage tissues. Dissected TGCs (DGC - E10 and E11), ectoplacental cone (EPC - E10 and E11), and E13, E15 and E17 placental samples. *Psg* expression is normalised to *Hprt* expression. (A) *Psg19* expression, (B) *Psg21* expression, (C) *Psg22* expression and (D) *Psg23* expression. (n=2).

gene transcription, were also dissected from E13, E15 and E17 pregnant timed-mated females. A study of *Psg* expression patterns in these tissues was undertaken, (Fig:3.14.). As in the previous expression quantification experiment, the expression of four *Psg*, *Psg19*, *Psg21*, *Psg22* and *Psg23* was investigated. Unlike in TSC and TGC differentiation, these four murine *Psg* exhibited distinctly different patterns of expression between tissue samples. *Psg22* displayed the greatest level of expression in E10 dissected TGCs, the expression of which is approximately 10^6 fold greater than *Psg21* and *Psg23*, and a 10^3 fold greater than *Psg19*. This high level of *Psg22* expression is converse to *Psg21* and *Psg23*, which showed low levels of *Psg* expression at this time-point in the development of the placenta. Unlike *Psg22* and *Psg19*, whose expression in dissected TGCs decreased from E10 to E11, expression of *Psg21* and *Psg23* increased from E10 to E11. Low levels of *Psg* transcription were also detected in dissected EPC; this may represent contamination from adherent TGC or, alternatively, the earliest manifestation of differentiating SpT from late E10 and E11. In the dissected EPC samples, *Psg22* was the only *Psg* of the four genes investigated to show a decrease in expression from E10 to E11. Expression of *Psg19*, *Psg21* and *Psg23* all increased from E10 to E11 in dissected EPC samples. *Psg* expression patterns differ in dissected placental samples also. *Psg22* expression decreased in placental tissue samples from E13 to E17. The same can be seen in *Psg19* expression although levels of expression are considerably lower than *Psg22*. The expression of *Psg21* and *Psg23* increases as the placenta develops from E13 to E17. Both *Psg21* and *Psg23* display the same expression profile and level of expression in these tissues. Comparatively, *Psg22* and *Psg19* show similar expression profiles, except in the EPC tissues. *Psg19* expression levels are similar to *Psg21* and *Psg23*. These corresponding expression patterns for these two gene sets, follows the phylogenetic relationships between *Psg21* and *Psg23*, and *Psg19* and *Psg22* (Fig:3.2.A).

3.1.9 Identification and Quantification of the *Psg22* splice variant expression

A novel *Psg22* splice variant was discovered when amplifying the Open Reading Frame (ORF) of *Psg22* to construct a *Psg22* expression vector to be used to produce purified recombinant Psg22 protein as described in materials and methods. Primers *Psg22* ORF F and R amplified two variants of *Psg22*, named hereafter as *Psg22* Long and *Psg22* Short. Amplicons of 1425 bp and 1069 bp were obtained when using E15 placental cDNA as template. RT-PCR products were cloned and sequenced, and returned sequences were compared (Fig:3.15A&B). The alternative splice variant of *Psg22* contained a deletion of the N1 (IgV-like) domain (*Psg22*Δ*N1*). This alternative splice variant had not been described previously and did not feature on any of the public databases. Both splice variants of *Psg22* are expressed in differentiated TSGs, with low levels of both transcripts being expressed in undifferentiated TSC (Fig:3.16.A). The discovery of this novel *Psg22* splice variant led to the investigation into the abundance of both of these variants' expression in trophoblastic tissues and cell lines. Splice variant transcript quantification was performed as described elsewhere [251], employing a dual insert plasmid containing specific distinguishable regions of both transcripts to construct a standard curve for qRT-PCR analysis. Relative quantification of each variant was performed as described in materials and methods.

Two TSC lines and their differentiated TGCs were used for *Psg22* splice variant expression analysis by qRT-PCR. The relative abundance of both *Psg22* variants in TSC and differentiated TGCs is shown (Fig:3.16.A). Both variants show the same expression patterns in these cell lines. Although, there is greater than a 100 fold difference in expression levels between the full length *Psg22*, (*Psg22* Long) and the truncated splice variants (*Psg22* Short) in undifferentiated TSC. There is an upregulation of both *Psg22* variant transcript expression, with a 10 fold difference in expression levels between *Psg22* Long and Short transcripts when TSC undergo differentiation towards the TGC fate.

A

Psg22 Long transcript:

ATGGAGGTATCCTCTGAGCTTCTCAGCAATGGGTGGACCTCCTGGCAAAGGGTTCTGCT
CACAGCCTCCCTCTTAACCTGCTGGCTCTTGCCCATCACTGCCGGAGTCACCATCGAAT
CCGTACCACCCAAATTGGTTGAAGGAGAAAATGTTCTTCTACGAGTGGACAATCTGCCA
GAGAATCTTCGAGTCTTTGTCTGGTATAGAGGGGTGACAGACATGAGCCTCGGAATTGC
ATTGTATTCACTTGACTATAGCACAAGTGTGACAGGACCTAAGCACAGTGGTAGAGAGA
CATTGTACAGAAACGGGTCCCTGTGGATCCAAAATGTCACCCGGGAAGACACAGGATAT
TACACTCTTCAAACCATAAGTAAAAATGGAAAAGTGGTATCAAATACATCCATATTCCT
TCAGGTGAACTCCTCTCTTTTCATCTGTGGGCGCCCTTCTCCACCTGCACTCCTCACTA
TTGAATCAGTGCCAGCCAGCGTTGCTGAAGGGGGAAGCGTTCTTCTCCTTGTCCACAGT
CTTCCAGATAATCTTCAATCGCTTCTCTGGTACAAAGGGTTGACTGTGTTTAACAAGGT
TGAGATTGCTCGGCACAGAACAGTCAAGAATTCAAGTGAAATGGGCCCTGCCTACAGCG
GTAGAGAGATAGTGTACAGCAATGGATCTCTGCTGCTCCAGAATGTCACCTGGGAAGAC
ACAGGATTCTACACCCTACAAATTGTGAACAGATATTGGAAAATGGAATTAGCACACAT
TCTTCAGGTGGACACCTCCCTTTCCTCGTGCTGTGACGATTTCAACTCTGTCCAACTGA
GGATCAATCCAGTGCCACCGCATGCTGCTGAAGGGGAAAGGGTTCTTCTCCAGGTCCAT
AATCTGCCAGAAGATGTGCAAACCTTTTTGTGGTACAAAGGCGTCTATAGCACTCAGAG
CTTTAAAATTACAGAGTATAGCATAGTGACAGAGTCTCTCATCAATGGCTATGCACACA
GTGGAAGAGAGATATTGTTCATCAATGGATCCCTGCTGCTCCAGGATGTCACTGAGAAA
GACTCTGGCTTCTACACACTAGTAACAATCGACAGCAATGTGAAAGTTGAAACAGCCCA
TGTGCAAGTCAATGTGAACAAGCTTGTGACACAGCCTGTCATGAGAGTCACGGACAGCA
CAGTTCGAATACAGGGCTCAGTGGTCTTCACTTGCTTCTCAGACAACACTGGGGTCTCC
ATCCGTTGGCTCTTCAACAATCAGAATCTGCAGCTCACAGAGAGGATGACCCTGTCCCC
ATCAAAGTGCCAACTCAGGATACATACTGTGAGGAAGGAGGATGCTGGAGAGTATCAAT
GTGAGGCCTTCAACCCAGTCAGCTCAAAGACCAGTCTCCCAGTCAGGCTGACTGTGATG
AATGAGTGA

B

Psg22 Short transcript:

ATGGAGGTATCCTCTGAGCTTCTCAGCAATGGGTGGACCTCCTGGCAAAGGGTTCTGCT
CACAGCCTCTCTTTTCATCTGTGGGCGCCCTTCTCCACCTGCACTCCTCACTATTGAAT
CAGTGCCAGCCAGCGTTGCTGAAGGGGGAAGCGTTCTTCTCCTTGTCCACAGTCTTCCA
GATAATCTTCAATCGCTTCTCTGGTACAAAGGGTTGACTGTGTTTAACAAGGTTGAGAT
TGCTCGGCACAGAACAGTCAAGAATTCAAGTGAAATGGGCCCTGCCTACAGCGGTAGAG
AGATAGTGTACAGCAATGGATCTCTGCTGCTCCAGAATGTCACCTGGGAAGACACAGGA
TTCTACACCCTACAAATTGTGAACAGATATTGGAAAATGGAATTAGCACACATTCTTCA
GGTGGACACCTCCCTTTCCTCGTGCTGTGACGATTTCAACTCTGTCCAACTGAGGATCA
ATCCAGTGCCACCGCATGCTGCTGAAGGGGAAAGGGTTCTTCTCCAGGTCCATAATCTG
CCAGAAGATGTGCAAACCTTTTTGTGGTACAAAGGCGTCTATAGCACTCAGAGCTTTAA
AATTACAGAGTATAGCATAGTGACAGAGTCTCTCATCAATGGCTATGCACACAGTGGAA
GAGAGATATTGTTCATCAATGGATCCCTGCTGCTCCAGGATGTCACTGAGAAAGACTCT
GGCTTCTACACACTAGTAACAATCGACAGCAATGTGAAAGTTGAAACAGCCCATGTGCA
AGTCAATGTGAACAAGCTTGTGACACAGCCTGTCATGAGAGTCACGGACAGCACAGTTC
GAATACAGGGCTCAGTGGTCTTCACTTGCTTCTCAGACAACACTGGGGTCTCCATCCGT
TGGCTCTTCAACAATCAGAATCTGCAGCTCACAGAGAGGATGACCCTGTCCCCATCAAA
GTGCCAACTCAGGATACATACTGTGAGGAAGGAGGATGCTGGAGAGTATCAATGTGAGG
CCTTCAACCCAGTCAGCTCAAAGACCAGTCTCCCAGTCAGGCTGACTGTGATGAATGAG
TGA

Figure 3.15: *Psg22* Full length (Long) transcript and *Psg22* (Short) transcript splice variant CDS sequences. (A) *Psg22* Full length (Long) transcript CDS sequence (1425 bp). Red nucleotide sequence indicates spliced sequence which contains the N1 domain. (B) *Psg22* Short splice variant CDS sequences (1065 bp).

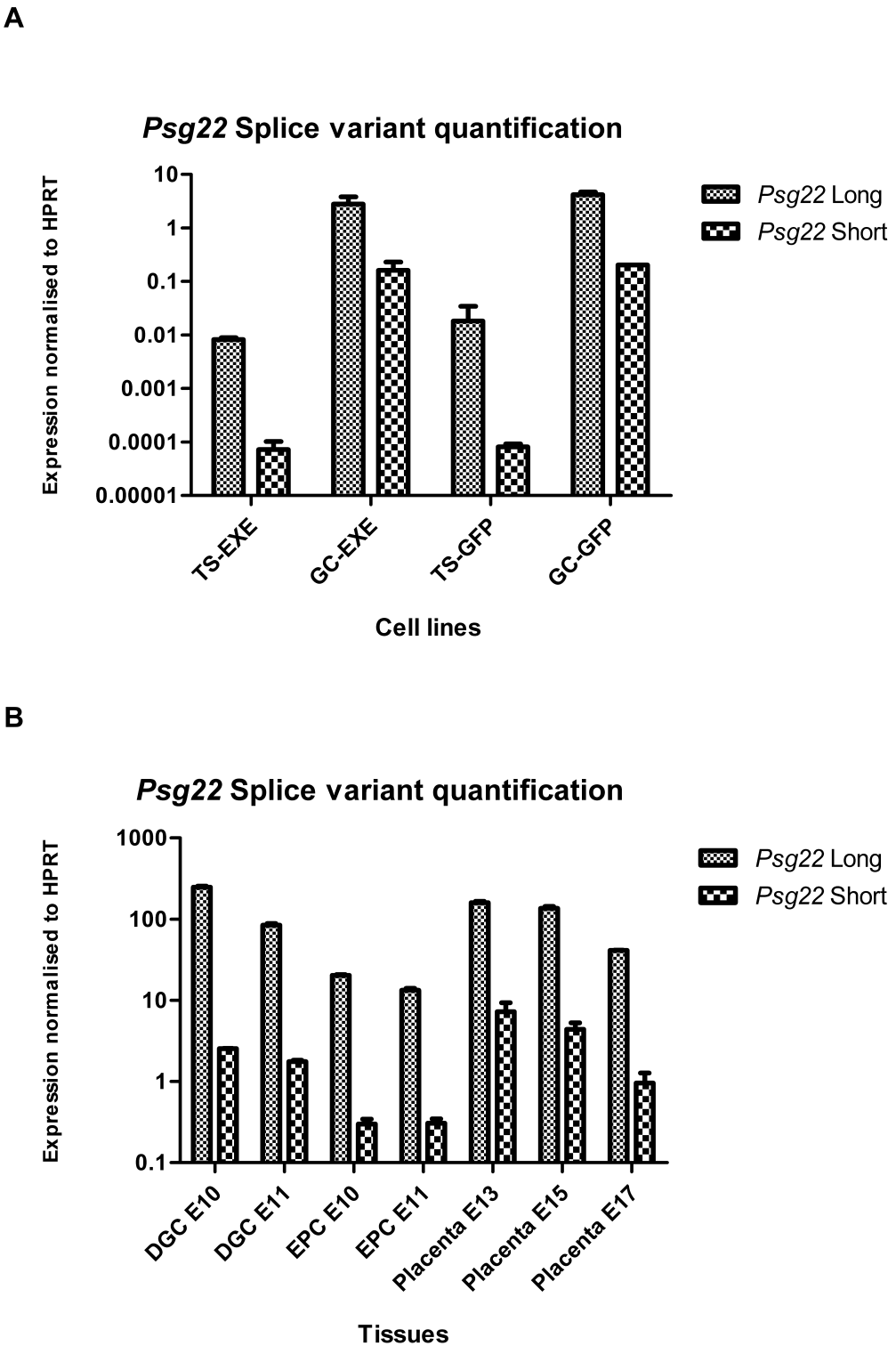


Figure 3.16: Relative quantification of *Psg22* Full length (Long) transcript and *Psg22*Δ*N1* (Short) transcript in (A) TSC lines (TS-EXE and TS-GFP) and 6 day differentiated TGCs. (n=3). (B) Dissected TGCs (DGC - E10 and E11), ectoplacental cone (EPC - E10 and E11), and E13, E15 and E17 Placental samples. (n=2).

The relative abundance of both variants in a number of trophoblast lineage tissues was then investigated. E10 and E11 dissected TGCs, E10 and E11 dissected EPC, and three time points (E13, E15, and E17) of full placental tissue were used as templates for *Psg22* splice variant expression analysis in qRT-PCR reactions. (Fig:3.16.B) shows that both splice variants possess the same expression profile as total *Psg22* in dissected TGCs and dissected EPC (Fig:3.14.C). *Psg22* splice variants are highly expressed in E10 dissected TGCs, and expression of both variants decrease as the placenta develops. It was found that full length *Psg22* has approximately 100 fold higher level of expression than *Psg22* Short in all tissues and time-points tested. The highest levels of the *Psg22* Short variant was found in E13 placental samples. Both variants have higher expression levels in earlier time points in all tissues investigated and expression decreases as embryonic development progresses. The discovery of a novel *Psg22* transcript variant and the fact that this transcript has a significantly different level of expression than the full length *Psg22*, poses the question of whether these two transcripts encode for proteins with the same function. I will address this question in the next chapter.

3.1.10 Investigating *Psg22* translation efficiency

Expression analysis in trophoblast tissues has shown that *Psg22* is highly expressed in dissected TGCs (Fig:3.8.C). Whether this highly expressed transcript correlates with a high level of translation was investigated in this section. High levels of transcription does not always mean that these transcripts are efficiently translated. Discrimination between actively translated and translationally silent mRNAs in the cell can be carried out using sucrose-gradient fractionation (polysome gradients), since this technique allows separation of free ribonucleoprotein particles (ribosome-free mRNA) from mRNAs bound to ribosomes (polysome-bound mRNA); thus ribosome loading of a transcript is a robust indicator of translation efficiency [253]. This technique allows for the determination of the fraction of a specific mRNA bound to ribosomes versus the fraction existing as free messenger ribonucleoprotein particles (mRNPs), thus giving an estimate of its translation efficiency. By comparing this parameter between different

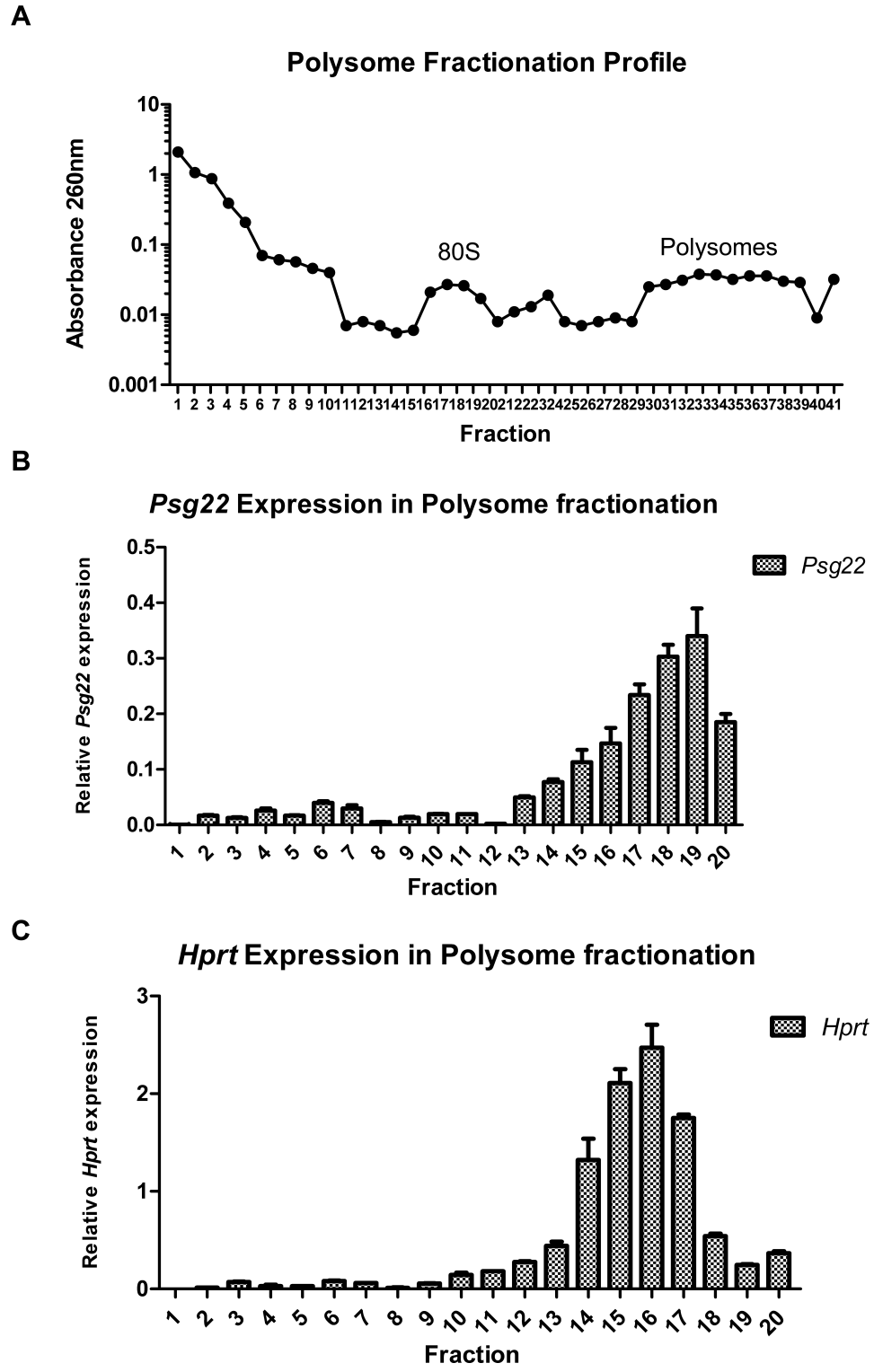


Figure 3.17: Investigating *Psg22* translation efficiency - Polysome fractionation of *Psg22* transcripts. Approximately 20 mg of lysed E10 dissected TGC tissue was used as template. (A) A260 nm readings of 40 fractions. (B) Quantification of *Psg22* transcripts in pooled fractions. *Psg22* transcripts are found in the polysome bound fractions indicating that *Psg22* is efficiently translated. (C) Quantification of *Hprt* transcripts in pooled fractions as a positive control.

stimulation or growth conditions, or between different cell types, it is possible to estimate the degree of translational control. For many mRNAs not all functional molecules are attached to ribosomes; some exist as free mRNPs. Technically, free mRNPs and polysome-bound mRNAs are separated by the principle of sedimentation velocity in a sucrose gradient. While free mRNAs will not enter the gradient, the migration of ribosome-bound transcripts is directly proportional to their loading with ribosomes, due to the increase in density of polysomes over free mRNPs. After the run, the gradient is fractionated and analysed [259].

20 mg of dissected TGC tissue was used as template for this technique, the tissue was pulverised into a powder under liquid nitrogen, before being lysed, (ribosomes were immobilised on their transcripts using 100 µg/ml cycloheximide present in the lysis buffer), cell debris removed, and the supernatant was extracted and carefully placed on a 11 ml 10-60% sucrose gradient as described previously [254]. Gradients were centrifuged for 3 hours at 38,000x g at 4°C, and 40 fractions were carefully collected from the top. Total mRNA in each fraction was determined using a A260/280 UV spectrometer. These UV spectrometer results were plotted to produce an RNA profile, (Fig:3.17.A). Free mRNPs can be seen in fractions 1-13, while fractions 16-20 contain 80S RNA. Polysome bound RNA can be detected in fractions 30 to 39, as depicted by the peak in the RNA profile. RNA was extracted and cDNA synthesised. Fractions containing *Psg22* transcripts were determined using qRT-PCR, with *Psg22* specific qRT-PCR primers. *Hprt* specific qRT-pCR primers were also used as a control for polysome loading. Primer sequences are listed (Table 2.6). *Psg22* transcripts are found in fractions 15 to 20, indicating that these transcripts are heavily loaded with ribosomes (Fig:3.17.B). Heavy ribosome loading of transcripts demonstrate that these transcripts are translated. The expression of *Hprt* in these fractions is shown in (Fig:3.17.C). *Hprt* transcripts can be found in fractions 14-18, indicating that these transcripts are also heavily loaded with ribosomes. Both *Psg22* and *Hprt* have the same polysome fractionation profile, indicating that both these transcripts are translated. The fact that *Psg22* transcripts have high levels of expression, and these results demonstrating that *Psg22* transcripts are associated with polysomes corroborates the

hypothesis of *Psg22* playing a role in TGCs.

3.2 Function and Regulation of murine *Psg22*

3.2.1 Introduction

The determination of the regulatory and functional properties of genes is central to genetical and biochemical research. To improve our understanding of murine *Psgs*, I investigated the mechanisms responsible for the regulation of *Psg22* and a possible functional role for Psg22 protein. I constructed expression vectors that express both *Psg22* variants. Employing a mammalian HEK cell expression system, both protein isoforms of Psg22 were produced, and purified using affinity chromatography. The function of these Psg22 proteins was examined, specifically in their ability to induce anti-inflammatory cytokines, such as TGFβ1 in human and mouse monocytic and macrophage cell lines. Two *Psg22* short-hairpin RNA (shRNA) vectors were constructed to attempt to knockdown *Psg22* expression *in vitro*.

I investigated the possible mechanisms that are responsible for the tissue-specific regulation of *Psg22*, and other *Psgs*. Transcriptional regulation can occur at both genetic and epigenetic levels. Genetic regulation is defined as a direct or indirect interaction between a gene and a transcription factor, and epigenetic regulation as altering DNA accessibility to transcription factors by chemical modification of chromatin [260]. The transcription factors that may bind to *Psg* promoters was analysed using transcription factor binding analysis software, and the relative frequencies of transcription factors that are implicated in TGC differentiation and PSG regulation was examined. The conformation of local chromatin in the regulatory regions of *Psg22* and *Psg23* was investigated to determine whether the chromatin surrounding these regulatory regions was in an open conformation. The mechanisms responsible for the relatively high expression of *Psg22* was not determined by this promoter analysis, and I hypothesised that an alternative regulatory mechanism is responsible. During the examination of the *Psg22* loci, I found an EST transcript (BY564540), that is located upstream of *Psg22*, which may be involved in the regulation of *Psg22* expression. I was able to identify three regions of sequence similiarity to this EST using online BLAST alignment software. Through the use of expression analysis

and primer walking RT-PCR, I was able to determine that this EST is expressed in a TGC-specific manner, is over 6 kb long and is expressed in an antisense manner to *Psg22*. Employing qRT-PCR I determined the expression patterns of this EST relative to *Psg22* expression, which demonstrated that these transcripts possess a concordant expression pattern in trophoblastic tissues and cell lines. Finally I investigated the local chromatin conformation associated with these antisense transcripts. I determined that the low-level expression of these transcripts is correlated with the modulation of local chromatin into an open conformation and with the relative high expression of *Psg22* in a cell-specific manner.

3.2.2 Psg22 protein production

To elucidate if both of the Psg22 protein isoforms share a common function, endotoxin-free purified recombinant Psg22 proteins were produced as described in materials and methods. The ORFs of both *Psg22* variants were cloned into the pQE-Trisystem-His-Strep-1 expression vector (Fig:3.18.B) using restriction endonuclease sites *Nco1* and *Pml1* incorporated into the primer set used to amplify the ORFs from E15 placental cDNA (Fig:3.18.A). Positive clones were verified by restriction digest band patterning and sequencing. Purified plasmid DNA for both variants was transfected into Freestyle™ 293-F cells as per protocol. The optimum time of maximal protein concentration post-transfection was discerned using various time-points post transfection. 1 ml of Freestyle™ 293 Expression Medium supernatant was removed from the transfected culture flasks at 12, 24, 36, 48, 60 and 72 hours post-transfection. Supernatant was centrifuged for 5 minutes at 1000 rpm to clear cell debris. 30 µl of supernatant from each time-point was tested on a Western immunoblot to assess optimum transfection times for each Psg22 protein isoform. Psg22 Long protein was detected at 12 hours post-transfection, whereas, Psg22 Short protein was detected at 24 hours post-transfection (Fig:3.18.C). A Rabbit anti-Poly6xHis antibody was used to detect the presence of these proteins, as the pQE-Trisystem-His-Strep-1 expression vector incorporates a Strep-His tag onto the C-terminus of the proteins to facilitate in purification of these proteins from medium supernatant via affinity

chromatography. I found that 72 hours post-transfection, for both isoforms, is the optimum time to harvest cell culture medium for protein purification. 20 µg purified human recombinant PSG1 was used as a positive control for the detection of His-tagged proteins.

To verify that the Psg22 proteins were being efficiently secreted into the cell culture medium and not retained in the cells, 30 µl of 72 hours cell culture supernatant, and 20 µg of 72 hours post-transfection cell lysate were run on a polyacrylamide gel, and using the rabbit anti-Poly6xHis antibody, I detected recombinant Psg22 in cell supernatant samples but not in the cell lysates, indicating that both isoforms of Psg22 are secreted (Fig:3.18.D). 250 ml cultures of Freestyle™ 293-F cells were cultured to a density of 1 x 10^6 cells/ml, and 72 hours post transfection, transfection medium was harvested. Psg22 proteins were batch-bound to Ni-agarose beads in the presence of 10 mM imidazole overnight at 4°C. Psg22 proteins were eluted using increasing concentrations of imidazole. Batch bound Psg protein medium was run through 10 ml endotoxin-free polypropylene columns, and isolated Psg22 bound beads were washed with 6 ml of wash buffer. Washing was complete when no protein was detected with UV spectrometer in wash flow through. Psg22 proteins were eluted with increasing concentrations of imidazole (50 mM, 200 mM, 300 mM and 500 mM). 4 x 1.5 ml fractions of each concentration was collected, and 30 µl of each fraction collected was run through a polyacrylimide gel and Psg22 proteins were detected using rabbit anti-Poly6xHis Ab (1:1000 dilution) (Fig:3.18. E). The western immunoblot demonstrates that the majority of Psg22 protein was eluted in the 200 mM imidazole fractions. The 200 mM and 300 mM imidazole fractions were pooled together, and added to 10 kDA cut-off protein spin columns, centrifuged at 3600 rpm until concentrated to 2 ml. Concentrated protein solutions were dialysed using dialysis cassettes in 1600 ml endotoxin-free PBS overnight at 4°C. A second round of dialysis was performed for another four hours in fresh endotoxin-free PBS. Purified protein was concentrated further to 500 µl, aliquoted, and stored in 0.5 ml Eppendorf tubes at -80°C. Protein concentration was determined using the extinction coefficient quantification method.

To determine the purity of these Psg22 proteins, 2 µg of each purified protein

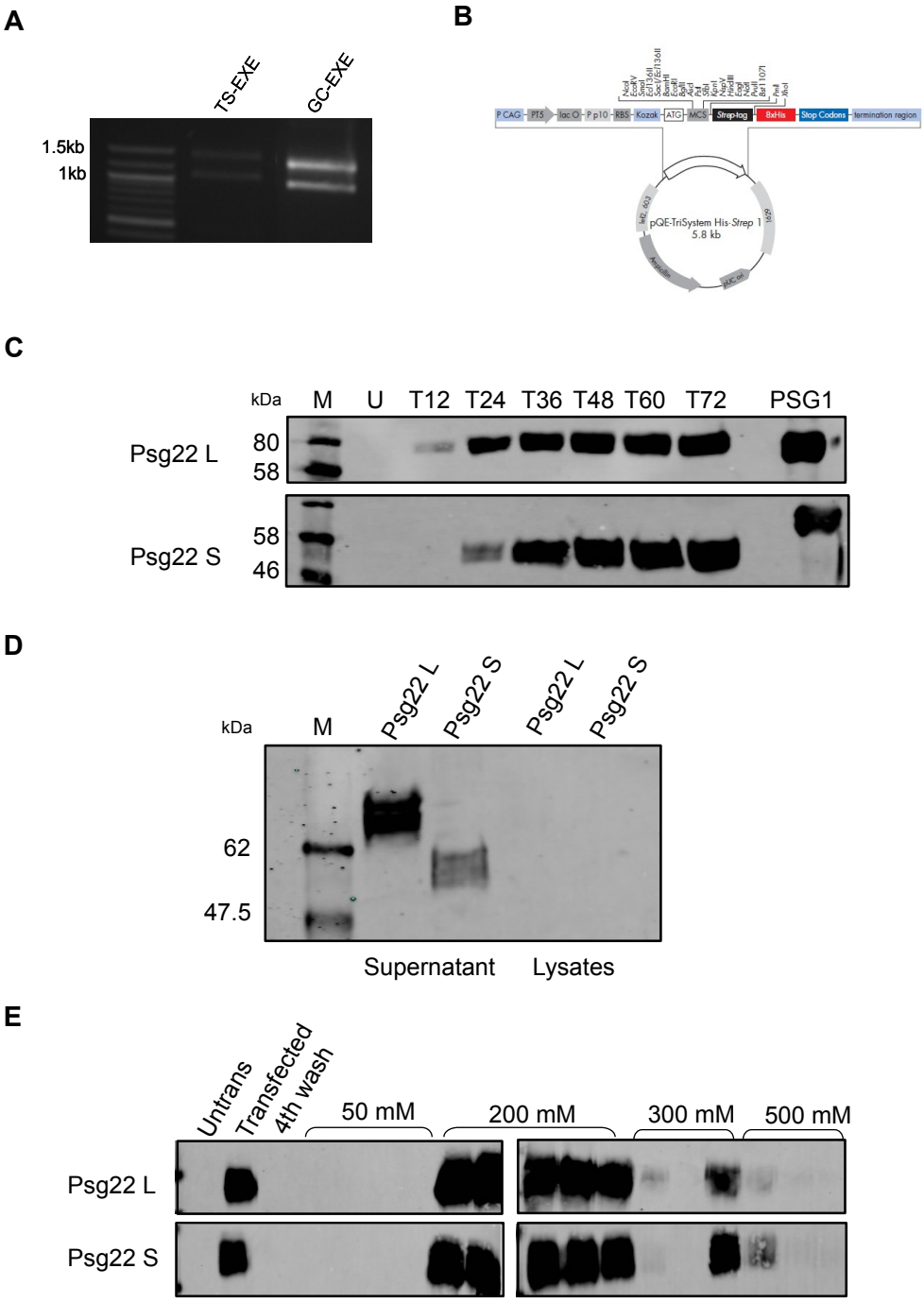


Figure 3.18: Optimisation of Psg22 protein production. (A) *Psg22* splice variant RT-PCR - TS-EXE and GC-EXE cDNA with *Psg22* ORF primers. (B) Schematic of PQE-Trisystem-His-Strep expression vector. (C) Western immunoblots of *Psg22* Long and Short test transfections. (D) Western immunoblot of 72 hours HEK293 post-transfection supernatant and lysates - rabbit anti-Poly6xHis Ab for both Psg22 protein isoforms. (E) Western immunoblot of imidazole elutions of purified Psg22 proteins - rabbit anti-Poly6xHis Ab.

was run through a polyacrylimide gel which and stained with Coomaisse brilliant blue dye to visualise the protein bands. Both proteins are present, with the Psg22 Long isoform transfection producing a very pure protein, although there were a few nonspecific protein bands present in the Psg22 Short protein preparation (Fig:3.19.A). As reported previously, mutated Psg proteins have been found to produce unknown higher unspecific bands [170], resulting in protein purity of approximately 90%. The *Psg22* Long transcript produces a protein of 55 kDa, which includes the addition of the Strep-His tag. Post-translational glycosylation of these proteins increases the molecular weight by approximately 30% [261]. This post-translational modification results in a full length protein of approximately 71 kDA. The *Psg22* Short transcript encodes for a protein of 42 kDa, including the Strep-His tag. The Psg22 Short isoform of this protein after post-translational modification has a molecular weight of 54 kDa. Fig:3.19.B shows a schematic of Psg22 protein isoforms domain predictions.

N-terminal sequencing was employed to determine the first 5 aa sequence of these two purified proteins (Alta-Biosciences, UK). It revealed that both proteins are cleaved at the predicted end of the Leader sequence (Fig:3.19.C), which is at position 34 aa for Psg22 Long, and at position 30 aa for Psg22 Short. The first 5 aa of the Psg22 Long isoform are **VTIES**, in comparison to the Psg22 Short isoform, which is **SPPAL**, resulting in a protein that is 115 amino acids shorter than the full length Psg22 protein. Both of these proteins are identical, with the exception of the IgV-like N1 deletion, present in the Psg22 Short protein. As previously discussed, the RGD-like motif that is located in the IgV-like N1 domain in all rodent and human PSGs, is implicated as a key motif involved in PSG functionality. The omission of the IgV-like N domain in the truncated Psg22 Short protein, may have detrimental effects on this Psg22 variants' function. I therefore tested whether these proteins share the same function, or have different functions.

Two rabbit polyclonal antibodies, anti-Psg22N1A and anti-Psg17N1, were kindly donated by G. Dveksler. Western immunoblotting was used to test the specificity of these antibodies. Recombinant PSG proteins (PSG1, PSG9, Psg22 Long and Short, Psg22N1A, & Psg17N1) were tested to check for cross-Psg reactivity,

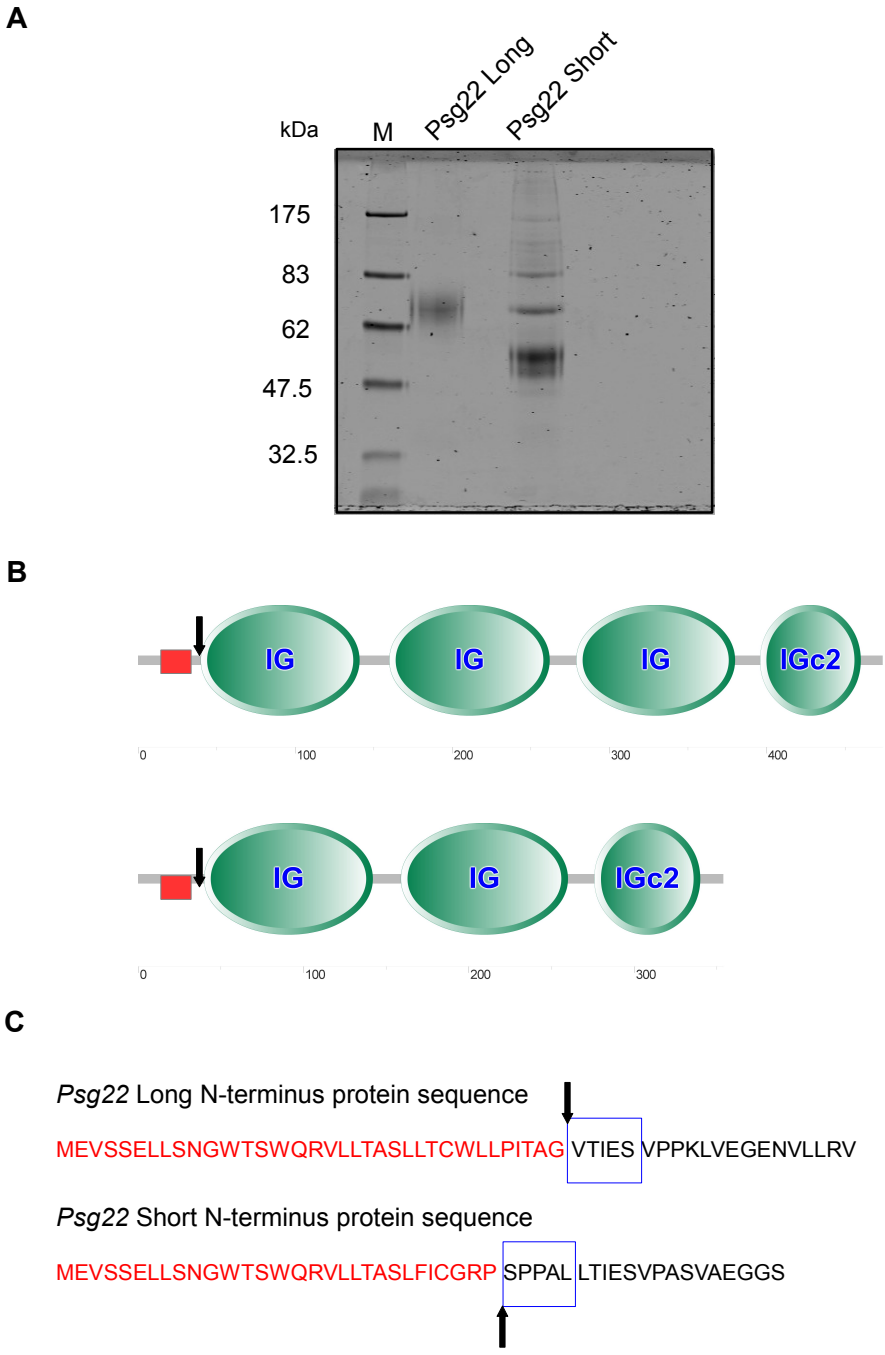


Figure 3.19: Psg22 protein purification and murine Psg antibody characterisation. (A) Coomaisse stain of 2 µg purified Psg22 Long and Short protein isoforms. (B) Schematic of Psg22 splice variant domain organisation. SMART (a Simple Modular Architecture Research Tool) output. (C) N-terminal sequencing of purified Psg22 proteins – leader sequence cleavage and first five amino acids sequenced (leader sequence denoted in red).

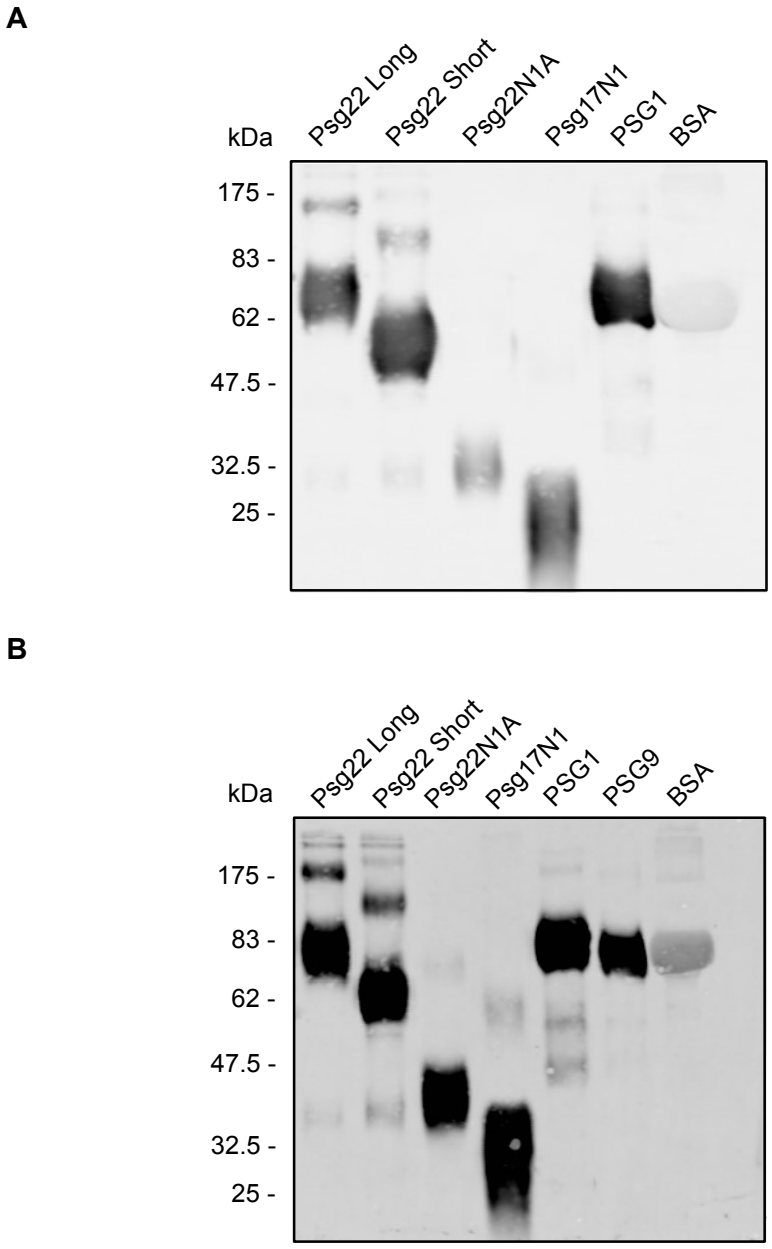


Figure 3.20: Polyclonal anti-Psg antibody characterisation. (A) polyclonal rabbit anti-Psg17N1 antibody characterisation (B) polyclonal rabbit anti-Psg22N1A antibody characterisation. 2 µg of each purified recombinant protein, including mouse Psg22 Long and Short isoforms, Psg22N1A, and Psg17N1, human PSG1, PSG9 and BSA standard, were used as samples. Both primary antibodies used at 1:800 dilution. Secondary antibody goat anti-rabbit IRDYE 680 (LI-COR) was used at 1:1000 dilution

and cross-species PSG reactivity. Both antibodies detects, both long and short isoforms of recombinant Psg22 protein, Psg22N1A, Psg17N1, and human PSG1 (Fig:3.20.A&B). The fact that these antibodies have cross-PSG and cross-species reactivity is unsurprising given that these antibodies are polyclonal, and PSGs are cloesly related. The fact that these antibodies cross-react with other murine Psgs, and possibly the related CEACAMs (which also posses a similar IgV-like N domain), renders them unsuitable for use in the specific detection of Psg22, and they were not used further in this study.

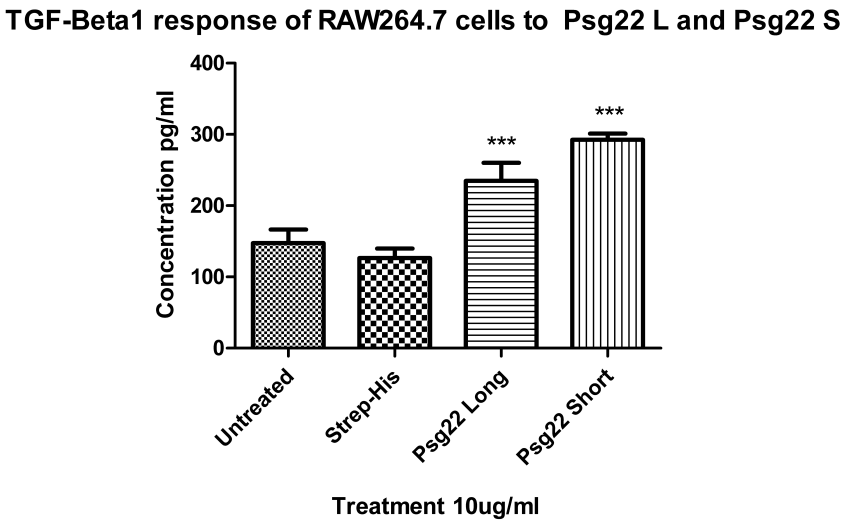
3.2.3 Psg22 induction of TGFβ1 - ELISA analysis

TGFβ1 has pleiotrophic effects in regulating T cells, B cells, and macrophages. TGFβ1 has been found to be produced by every leukocyte lineage, including lymphocytes, macrophages, and dendritic cells, and its expression serves in both autocrine and paracrine modes to control the differentiation, proliferation, and state of activation of these immune cells. TGFβ1 has been implicated in immuno-suppression, and it has been shown that the administration of TGFβ1 suppresses symptoms of certain experimentally induced autoimmune diseases whereas the administration of anti-TGFβ1 antibodies exacerbates these conditions [262]. It has also been shown that TGFβ1 exerts systemic immune suppression and inhibits host immunosurveillance [263]. TGFβ1 is a proangiogenic factor that plays an important role in the development of the fetoplacental capillary system during implantation [264]. TGFβ1 has multiple roles during pregnancy, including regulation of extravillous trophoblast migration and proliferation and regulation of NK cell function [170]. It has been previously described that murine Psg proteins (including a truncated Psg22-N1-A protein) induce TGFβ1 in monocyte and macrophage cell lines [170, 265, 165]. To assess whether the two Psg22 protein isoforms that have been produced in this study share a common function, I tested their ability to induce TGFβ1 in a murine RAW246.7 macrophage cell line and in a human THP-1 monocytic cell line. Induction of TGFβ1 was measured using an eBiosciences Ready-Steady-Go TGFβ1 ELISA. The RAW246.7 and THP-1 cell lines were maintained as described in materials and methods. Cells were

plated in triplicate for each treatment in 24 well plates and incubated in a 37°C humidified incubator with 5% CO_2. Raw246.7 cells and THP-1 cells were seeded at a density of 1 x 10^6 cells/ml per well. Cells were treated with 10 µg/ml Psg22 Long and 10 µg/ml Psg22 Short on the following day in 300 µl of fresh media for 24 hours. Cells were also treated with 10 µg/ml recombinant PSG1 protein as positive control, and 10 µg/ml Strep-His peptide as a negative control. The Strep-His peptide (WSHPQFEKLEHHHHHHHH) (Eurogentec, Belgium) was used as a control for the Strep-His tag introduced to the C-terminus of the proteins expressed from the pQE-Trisystem-His-Strep-1 expression vector. This ensured that the tag was not responsible for TGFβ1 expression. After treatments, the supernatants were collected and centrifuged at 3000 rpm for 5 minutes to remove cell debris. The supernatants were activated as per protocol as this sandwich ELISA recognizes the mature/active form of TGFβ1. Samples (but not standards) were acid-treated and then neutralized to activate the latent TGFβ1 to the immunoreactive form.

In murine RAW246.7 cells (Fig:3.21.A), there is a higher induction of TGFβ1 from the Psg22 Short treatments than with the Psg22 Long protein treatments. Psg22 Short treatments induced TGFβ1 to levels of approximately 290 pg/ml, which is consistent with previous reports of TGFβ1 induction in RAW246.7 cells by Psg23N1A [170]. In THP-1 cells (Fig:3.21.B), TGFβ1 induction is much higher than in RAW246.7 cells, as is consistent with previous reports [170]. Induced levels of TGFβ1 by the full length Psg22 Long protein reach levels of nearly 3000 pg/ml, in contrast to TGFβ1 levels of approximately 6700 pg/ml are induced by the Psg22 Short protein treatments. The Psg22 Short protein treatments result in over a two fold the induction of TGFβ1 than the full length protein. This is due to these proteins not being used in equimolar concentrations, resulting in a higher dosage of Psg22 Short than Psg22 Long protein. Both Psg22 proteins have induced TGFβ1 significantly more than control treatments ($P \leq 0.001$) in RAW246.7 and THP-1 cell lines (Fig:3.19.A&B). Both of these Psg22 protein isoforms share the ability to induce TGFβ1, despite the difference in levels of TGFβ1 upregulation. This demonstrates that regardless of the fact that the Psg22 Short protein possesses a N1 domain deletion, a region which contains the 'RGD'-like

A



B

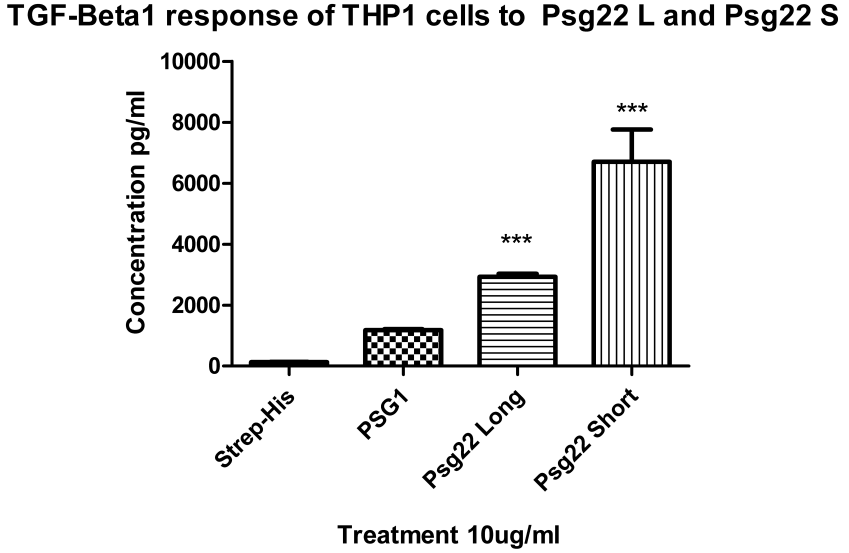


Figure 3.21: Induction of TGFβ1 by recombinant Psg22 proteins. (A) Mouse macrophage RAW264.6 cells were treated with 10 µg/ml of Psg22 Long and Short recombinant proteins for 24 hours in 24 well plate (n=3). Strep-His peptide used as a tag control. (B) Human THP1 cells were treated for 24 hours with 10 µg/ml Psg22 Long and Short protein isoforms. 10 µg/ml of human recombinant PSG1 protein used as control. 24 hours post treatments, cell medium supernatant was collected and induction of TGFβ1 was measured by ELISA. (n=3). Data was subjected statistical analyses using a One Way ANOVA and Bonferroni's Multiple Comparison Post-test. (***, $P \leq 0.001$)

motif and was previously implicated as playing a role in *Psg* functionality, both these proteins are able to induce TGFβ1. This leads to the conclusion that it is not the *Psg* N1 domain that is exclusively responsible for the induction of TGFβ1.

3.2.4 *Psg22* shRNA vector testing *in vitro*

Current approaches to study gene function, such as gene targeting via homologous recombination in murine embryonic stem (ES) cells has been the main approach used to investigate mammalian gene function *in vivo*. Even though there has been recent advances in this technology, it still remains a time-consuming, expensive and laborious method, that cannot be applied to human tissues. Important advances in RNA interference (RNAi) technology has produced a less-time consuming method for producing knockdown of gene expression to investigate gene function in a number of organisms using plasmid-based RNAi to stably silence gene expression [266, 267]. An RNA polymerase III promoter is used to transcribe a short stretch of inverted DNA sequence, forming a short hairpin RNA (shRNA) that is processed by Dicer to generate siRNAs [250]. The Cre-Lox conditional pSico Reverse (pSicoR) vector used in this research was generated by modification of the pLL3.7 vector, that expresses RNAi inducing shRNAs under the control of the U6 promoter [250]. The U6 promoter has been widely used to drive the expression of shRNAs and a U6-based lentiviral vector for the generation of transgenic mice has been recently described [249]. This vector was engineered to co-express enhanced green fluorescent protein (EGFP) as a reporter gene to aid in assessing transfection efficiency. This pSicoR vector allows constitutive shRNA expression, which can be terminated by a Cre mediated recombination event [250].

To assess *Psg22* function *in vivo*, two *Psg22* shRNA vectors were constructed and the knockdown of *Psg22* expression was performed *in vitro* using TGC lines as a source of endogenous *Psg22* expression. Oligos that target *Psg22* were generated using the PSICOOLIGIOMAKER1.5 software programme available from the Jacks' Lab (http://web.mit.edu/jacks-lab/protocols/pSico.html). Two separate oligos (*Psg22* shRNA construct 1 and 2) were selected based on predicted targeting by

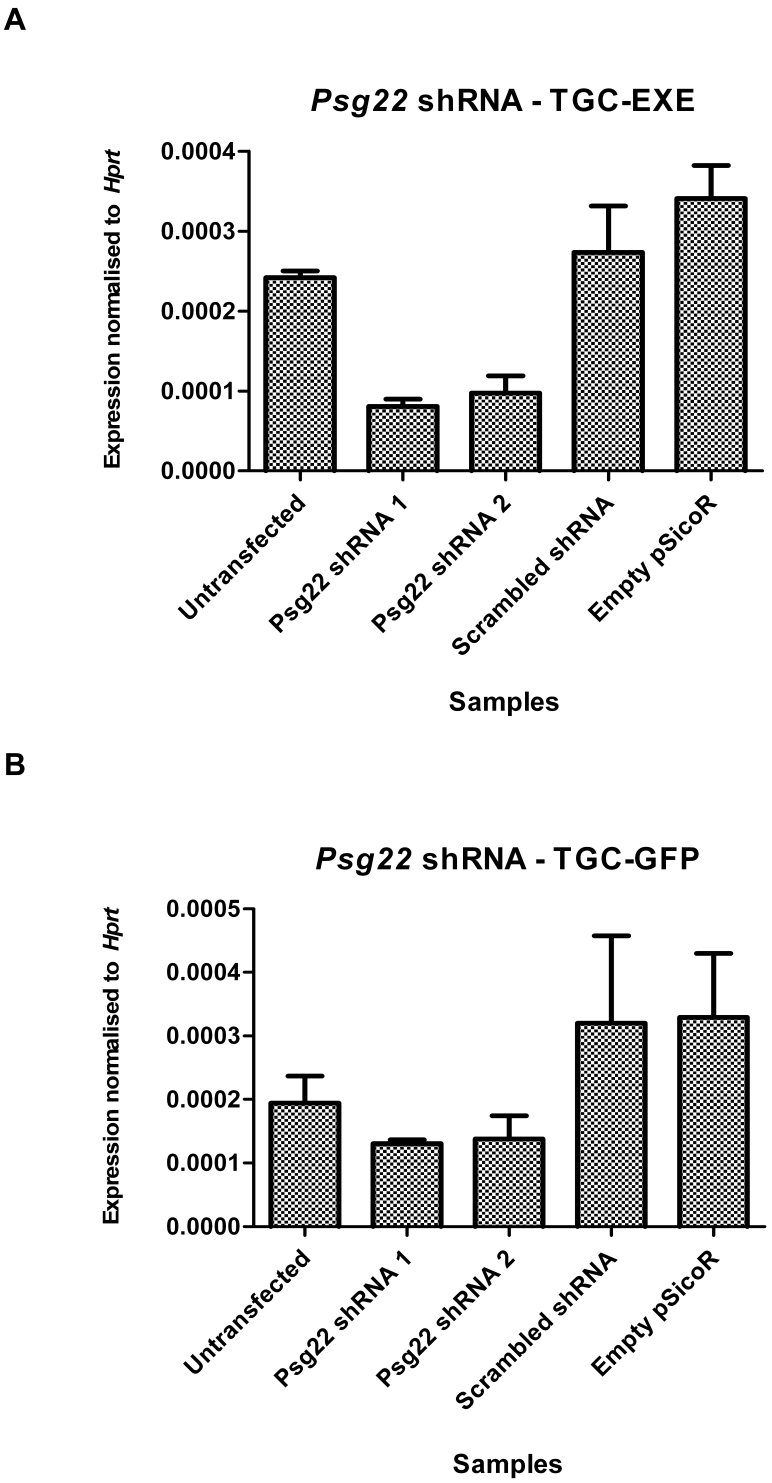


Figure 3.22: *Psg22* shRNA construct knockdown of *Psg22* expression in TGC lines. Two *Psg22* shRNA constructs, *Psg22* shRNA 1 and 2, scrambled shRNA and empty pSicoR vector controls. (A) *Psg22* shRNA knockdown of *Psg22* expression in GC-EXE cells. (B) *Psg22* shRNA knockdown of *Psg22* expression in GC-GFP cells. (n=1), best replicate of three independant experiments. *Psg22* expression normalised to *Hprt*.

PSICOOLIGOMAKER1.5. These oligos were designed to target both splice variants of *Psg22*. Oligos were ordered from MWG Eurofins (Germany). Sense and Antisense oligos were annealed as per Jacks' lab protocol, and cloned into the pSicoR expression vector as described in materials and methods. Completed *Psg22* shRNA pSicoR vectors were sent to GATC (Germany) for sequencing to confirm successfully cloned vectors. *Psg22* shRNA vector clones with correct sequences were then tested *in vitro* in terminally differentiated TGC lines. To transfect these shRNA vectors efficiently, two TSC (TS-EXE and TS-GFP) lines were seeded at a 80% confluency. Undifferentiated TSC were transfected using Lipofectamine2000 as per protocol in serum free medium. Six hours post-transfection, the serum free medium was replaced with TSC medium, to induce differentiation to a TGC fate. These cells were grown in TS medium for 6 days, producing a population of cells that contain a majority of TGC. These cells were then harvested, RNA was extracted, and cDNA was synthesised as per protocols. Using relative qRT-PCR, the extent of *Psg22* expression being knocked down by the *Psg22* shRNA constructs was assessed. Results were described as mean *Psg22* expression relative to mean *Hprt* expression. Primers used for qRT-PCR reactions are described (Table 2.6.). Three biological replicates of each cell line were evaluated, using three technical qRT-PCR replicates.

Untransfected TGCs, TGCs transfected with empty pSicoR vector, and TGCs transfected with an off-target shRNA construct were used as a control. In both TS-EXE and TS-GFP TSC lines, the empty pSicoR and off-target shRNA pSicoR control constructs had no affect on *Psg22* expression, which indicates that there is no unspecific knockdown of *Psg22* gene expression as a result of the pSicoR vector backbone. In TS-EXE cells the *Psg22* shRNA construct 1, produced the greatest knockdown in *Psg22* expression, with *Psg22* shRNA construct 2 producing a slightly less efficient knockdown (Fig:3.22.A). A similar knockdown of *Psg22*expression is found in TS-GFP cells transfected with these constructs. As in TS-EXE, the *Psg22* shRNA construct 1, produced a slightly better knockdown of *Psg22* expression than with *Psg22* shRNA construct 2 in TS-GFP cells (Fig:3.22.B). These results demonstrate a knockdown of *Psg22* transcript *in vitro* using *Psg22* shRNA vectors. This *in vitro*

testing demonstrates that these vectors could be used to knock down *Psg22* expression *in vivo*.

3.2.5 Investigation of murine *Psg* Promoters

As stated previously, human *PSGs* do not have conventional promoters, as promoters of human *PSG* genes are highly homologous and lack any obvious TATA-box, typical initiator elements, or large GC-rich sequences [202, 195]. I investigated whether murine *Psgs* possess similar regulatory promoter regions as human *PSGs*, and what mechanisms control the regulation of *Psg* transcription. The genes associated with regulation of human and murine *PSGs*, cell lines which have been used to demonstrate regulation, and published literature citations are listed (Table 1.5.). A database of all murine *Psg* promoter sequences was compiled. I analysed a 2 kb in length region as there was no obvious core promoter for murine *Psgs*, and the regions that could be responsible for regulation of *Psg* expression may lie within this extended 2 kb promoter region. The length of this 2 kb extended regulatory region was also chosen as it would allow for analysis of region-specific deletions in later experiments. This 2 kb extended regulatory region spans from -2000 bp 5' of each *Psg* to the base before the translational start site (ATG) designated (-1). I chose to include regions that span up to the ATG, as some *Psgs* contain a conserved regulatory region with human *PSGs* which is located approximately 180 bp upstream of the ATG site but lies inside the TSS in the 5'UTR. In all mouse *Psgs*, the ATG codon is approximately -200 bp downstream of the TSS located in exon 1. This 2 kb region upstream of the translational start site of all 17 *Psg* were analysed. As with the human *PSGs*, I was unable to find an obvious TATA box, or GC-rich regions. The homology of these *Psg* regulatory regions was analysed. These 2 kb regions were aligned using ClustalW and a neighbour-joined pairwise comparison phylogenetic tree was constructed as described previously. The regulatory regions of the murine *Psg* family showed homology of between 49 - 92%.

To investigate the transcriptional and regulational architecture present on murine *Psg* promoters, a database of the putative transcription factor binding sites implicated in TGC differentiation located on these 2 kb *Psg* extended promoters was

compiled, (Table 3.1.). These extended 2 kb regulatory regions were analysed using the MatInspector programme (Genomatix Software Suite, Germany) which identified putative transcription factor binding sites and the frequency of these binding sites for each transcription factor on individual *Psg* promoters. This analysis was performed using the MatBase database and the associated MatInspector algorithm implementing the optimum-threshold default parameter. MatInspector reduces the signal-to-noise levels associated with putative transcription factor binding analysis by limiting the number of predicted sites reported and only showing the highest-scoring matrix match per transcription factor family in the query sequence.

Although there were numerous putative transcription factor binding sites identified on the 17 murine *Psg* regulatory regions, this investigation centred on transcription factors involved specifically in TGC differentiation and known regulators of *PSG* expression. The transcription factor binding analysis focused on 15 transcription factors that have been previously implicated in either TGC differentiation or human *PSG* regulation (Table 3.1.). The roles of these transcription factors in TGC differentiation and the associated published literature regarding these TGC related transcription factors are highlighted (Table 1.2.). The results of this *Psg* extended regulatory region analysis has revealed a variety of transcription factors binding to different *Psg* regulatory regions at different locations and with different frequencies, which may explain the differences in individual *Psg* expression regulation.

I was especially interested in transcription factors that bind to the *Psg22* promoter that distinguishes this promoter from the rest of the murine *Psg* family, which may give an indication of the mechanisms which are responsible for the increased expression of *Psg22*. The only transcription factor that binds to *Psg22* that does not bind to the other *Psg* promoters is *FoxD3*. *FoxD3* is a member of the forkhead transcription factor family and has been implicated in the suppression of TGC differentiation [28, 87, 88, 89]. *FoxD3* is generally considered to be a transcriptional repressor and to be involved in the maintenance of pluripotency. However, *FoxD3* can also function as a transcriptional activator [268], and additional roles for *FOXD3* are

Table 3.1: *Psg* promoter Transcription Factor Binding Site analysis

TF	*Psg16*	*Psg17*	*Psg18*	*Psg19*	*Psg20*	*Psg21*	*Psg22*	*Psg23*	*Psg24*	*Psg25*	*Psg26*	*Psg27*	*Psg28*	*Psg29*	*Psg30*	*Psg31*	*Psg32*
AP-2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Hand1	3	0	0	1	0	1	1	0	2	1	0	0	0	2	0	0	1
Tead4	4	1	1	2	3	2	2	4	0	1	1	3	2	2	3	3	3
Cdx2	4	1	3	0	1	2	1	3	1	2	5	1	1	1	2	2	3
Gata2	4	3	5	3	3	0	2	2	2	2	6	4	2	3	5	5	4
Stat3	4	3	3	1	2	1	2	1	1	3	3	1	3	2	0	0	1
Ik3	0	3	1	2	1	1	3	2	2	1	2	2	3	2	2	1	1
FoxD3	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
NeuroD	1	0	1	1	1	1	2	2	0	0	2	1	1	0	0	0	0
Gcm1	1	2	3	2	3	4	2	1	8	2	6	2	4	3	4	8	5
Klf4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Klf6	1	2	0	1	0	0	1	0	1	0	0	0	0	2	0	0	0
SP1	1	3	3	3	2	0	1	0	1	1	1	2	1	3	1	2	4
FXRE	0	2	0	0	1	2	1	1	1	0	2	2	0	0	1	1	0
RxR	2	4	3	6	3	8	4	2	2	5	6	2	5	3	2	2	3

emerging particularly with regard to the differentiation of migratory cell phenotypes. Putative *FoxD3* transcription factor binding sites are located at the end of the analysed 2 kb promoter length (-1931 nt), far from the TSS and the second binding site is located at (-254 nt). To date there has been no evidence of *FoxD3* in the role of *PSG* regulation. Reporter construct assays are needed to elucidate if this transcription factor is involved in the suppression or activation of *Psg22* transcription. These stand-alone putative *Psg22* promoter *FoxD3* binding sites are intriguing, especially as there are no other *FoxD3* binding sites in a 2 kb region spanning all 16 other murine *Psgs*.

Lopez-Diaz *et al*, 2007, reported that the minimal promoter region of all *PSG* genes contains a putative Retinoic Acid Responsive Element (RARE) and that mutations at specific nucleotides within the RARE motif inhibits both RXRα-DNA interactions and RXRα transcriptional activation of *PSG5* promoter [200]. I investigated whether murine *Psgs* possessed this overlapping regulatory SP1-RARE site, using MatInspector transcription factor binding analysis software. 15 of the murine *Psg* possess a putatuive SP1 binding site, but only four out of 17 murine *Psgs* possess this overlapping SP1-RARE site in the CPE region. *Psg17, Psg19, Psg20* and *Psg26* all possess overlapping SP1-RXRα sites. Interestingly *Psg22* does not possess this overlapping dual transcription factor site, but only contains an SP1 site. Also of note, if the Guanine (G) base located at -35 on the *Psg22* 2 kb TSS upstream region, is mutated to a Cytosine (C), the RXRα site is reintroduced when analysed using the Transcription factor Binding Site software. I found that there are 4 other putative RxR binding sites along this 2 kb TSS upstream region of *Psg22*, and all 17 murine *Psgs* possess at least two RxR binding sites.

The *AP-2* γ or *Tfap2c* transcription factor, which is involved in TGC differentiation [78, 8], was found only to have one putative binding site on only one murine *Psg*, *Psg20*, suggesting that it does not have a role in *Psg22* transcriptional regulation. The Kruppel-like factor 4 (*KLF4*) which is implicated in the regulation of human *PSG5* transcription has no putative binding sites on any murine *Psg* promoter, indicating that it in not involved in directly binding to murine *Psg* DNA regulatory sequences. Its family member, *KLF6*, which also plays a role in *PSG5* regulation,

has binding sites on only six of 17 murine *Psg* promoters, being *Psg16, Psg17, Psg19, Psg22, Psg24* and *Psg29*. The *Psg22* promoter has two bHLH *Hand1* transcription factor binding sites, which is comparable to the frequency of *Hand1* sites on the other 7 *Psg* promoters which were found to posses this transcription factor. Also *Psg22* has two *NeuroD1* binding sites, which is comparable to the number of *NeuroD1* sites present on the 9 other *Psg* that contain these putative binding sites. *Stat3* has two putative binding sites on *Psg22* promoter, in contrast *Psg16* has four sites, which is the most putative binding sites in the 15 *Psg* with *Stat3* sites. *Tead4* sites can be found on all murine *Psg*, with the exception of *Psg24*. There are three *Ik3* binding sites on the *Psg22* promoter, which is the most sites presents on all murine *Psg* promoters except *Psg16*. The *Gata2/3* transcription factor is also present on all murine *Psg* promoters, with the exception of *Psg21*. There are two *Gata2/3* sites on the *Psg22* regulatory region. *Cdx2* is present on all murine *Psg*, except *Psg19*. The *Cdx2* transcription factor has 3 sites on the *Psg22* promoter, although, *Psg26* has 5 putative sites.

The only transcription factor, which is involved in TGC differentiation and is well represented in all mouse *Psg* promoters is *Gcm1*. *Psg22* only has two *Gcm1* binding sites, with *Psg24* and *Psg31* both containing 8 *Gcm1* regulatory regions. It is interesting that *Psg24* and *Psg31* share a common number of these regulatory sites, as they share the domain expansion of internal N domains, although their regulatory regions do not branch together on the phylogenetic tree. This suggests that *Gcm1* has a potential role in regulating all murine *Psg*, as there is a conservation of these sites amongst murine *Psg*. Further *in vitro* reporter construct analysis involving these individual transcription factors need to be employed to discern whether this *in silico* analysis has yielded transcription factor candidates that regulate murine *Psg* transcription.

Murine *Psg* 2 kb regulatory regions (Fig:3.23.A), do not follow the phlyogenetic relationships that is evident between the coding sequences of the *Psg* genes in (Fig:3.2.A). *Psg21* and *Psg23* 2 kb regulatory regions are highly related, as are *Psg30* and *Psg31* regulatory regions. *Psg26* and *Psg28* are also highly related. This phylogenetic tree reveals that the *Psg22* 2 kb regulatory region is located on

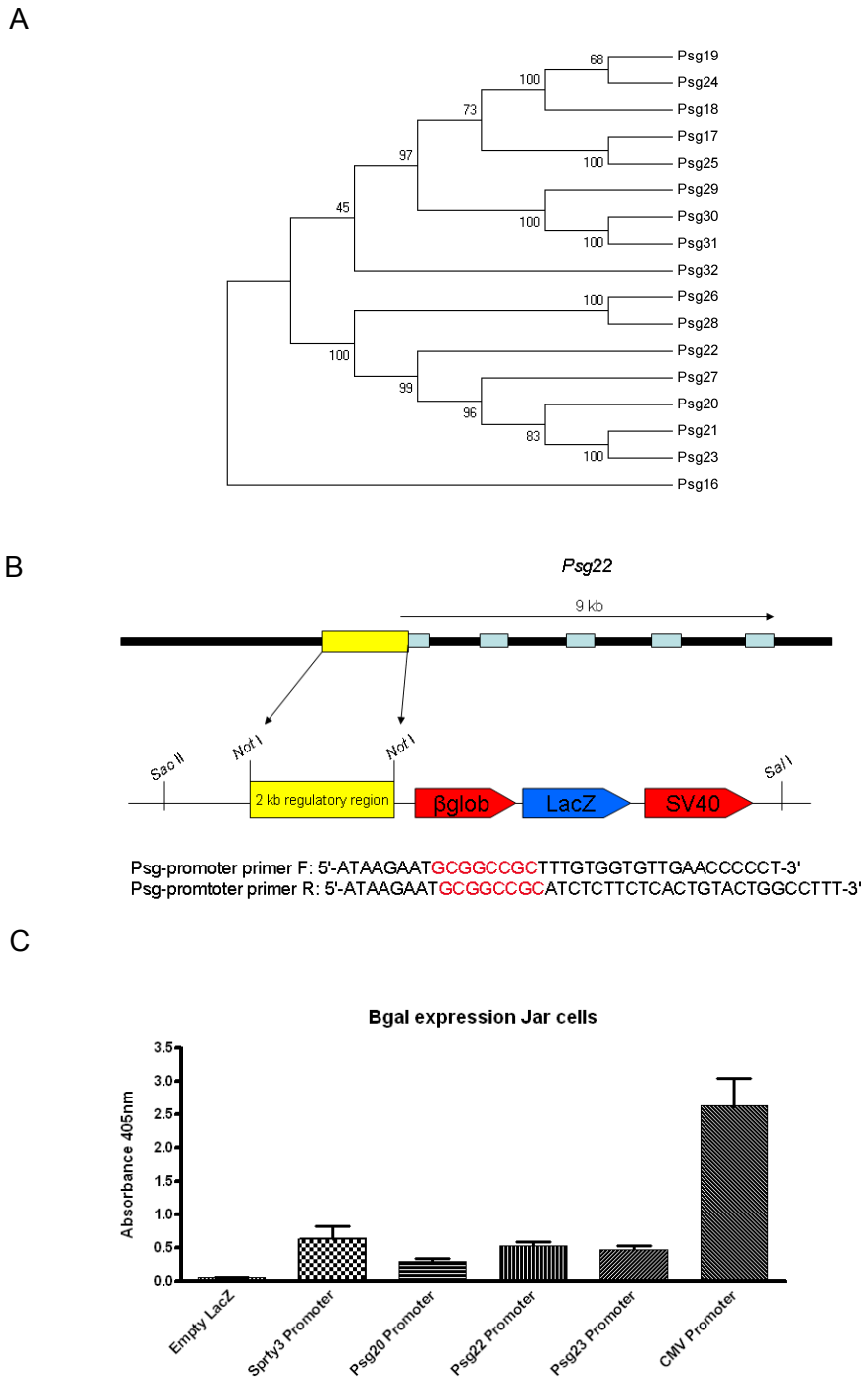


Figure 3.23: Investigating and quantifying *Psg* promoter activity. Promoter activity was quantified by induction of β-Galactosidase using Pierce Mammalian β-Galactosidase Assay Kit in Human choriocarcinoma JAR cell line. (A) *Psg* 2 kb upstream region neighbour-joined phylogenetic tree. (B) Schematic of *Psg* 2 kb upstream region inserted into *NotI* sites in LacZ expression vector. (C) β-Galactosidase quantification of promoter constructs in JAR cell line (n=3).

subbranch of it own, and its closest relatives are *Psg27, Psg20, Psg21* and *Psg23*. It is interesting to note that *Psg19* and *Psg22* 2 kb regions are quite different, even though these two genes' coding sequences cluster together on the same phylogenetic branch. This difference in 2 kb regulatory region similarity may be the reason why *Psg22* has a higher expression level than *Psg19* which is the closest relative of *Psg22*. Also of note is the location of the *Psg16* upstream region on this phylogenetic tree, which does not branch with any other *Psg* family member.

To investigate the promoter activity of murine *Psgs*, a quantitative LacZ expression assay was undertaken to assess the activity of *Psg20, Psg22* and *Psg23* 2 kb regulatory regions. *Psg20, Psg22* and *Psg23* 2 kb upstream regions were cloned individually into LacZ expression vectors. Cloned *Psg* 2kb upstream regions were sequence verified (GATC, Germany) and correctly engineered LacZ constructs were transfected into the human choriocarcinoma Jar cell line. Jar cells were maintained for 48 hours post transfection, and the resulting LacZ expression was measured using the Pierce ThermoScientific Mammalian β-Galactosidase Assay Kit as per manufacturers instructions. The schematic of the LacZ vector used and restriction sites employed in the cloning of the 2 kb *Psg* regulatory regions are shown (Fig:3.23.B). This LacZ expression vector contains a partial human β-globulin promoter linked to a LacZ gene coding region, and LacZ expression is driven by the region of DNA that is cloned between the two *NotI* restriction sites. β-Galactosidase activity is measured simply by colourimetric quantification.

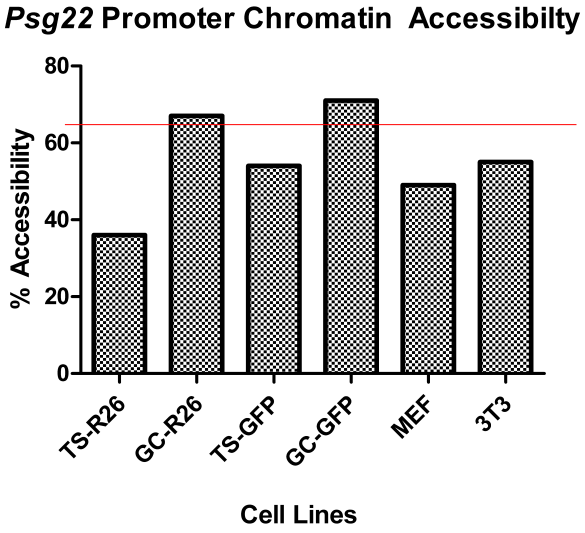
The quantified LacZ activity associated with each *Psg* regulatory region tested is shown (Fig:3.23.C). An empty LacZ vector was used as a negative control, and as can be seen, confers no promoter activity. A pCMV-SPORT-βgal LacZ construct was used as a positive control and gives the highest induction of LacZ in transfected Jar cells. This is due to the presence of the strong CMV promoter driving LacZ expression in this cell line. The Sprouty3 promoter positive control induced LacZ expression at slightly higher levels than the *Psg* 2 kb regulatory regions. The *Psg22* 2 kb region induced the highest level of LacZ expression of the three *Psg* regions analysed, although the difference in promoter activity between these three *Psg* 2 kb regions was

marginal. *Psg23* induced LacZ at a slightly lower level, and *Psg20* was found to induce the lowest levels of LacZ of the *Psg* 2 kb upstream regions tested. This result shows that the three *Psgs* upstream regulatory regions tested, have a low level of promoter activity in JAR cells.

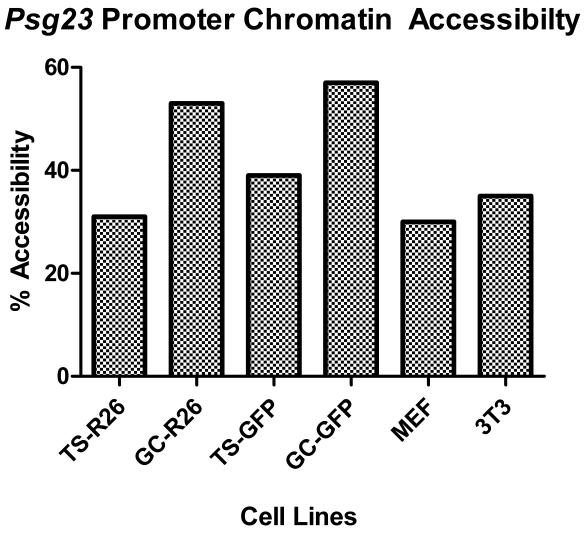
3.2.6 Investigation of chromatin structure and accessibility in *Psg* promoters

I investigated whether the 2 kb upstream region of *Psg22* possess an open chromatin conformation associated with *Psg22* transcription in TGC. I employed the EpiQ chromatin analysis kit (Bio-Rad) to assess the conformation of chromatin in cultured cells. The EpiQ kit quantifies the impact of epigenetic events, such as DNA methylation and histone modification, on gene expression regulation through chromatin state changes. This assay is based on the principle of *in situ* chromatin digestion, genomic DNA purification, and qRT-PCR to determine the chromatin environment of targeted regions of the genome. It can discriminate open, actively transcribed chromatin regions from closed, transcriptionally silent regions. Two TSC lines (TS-R26 and TS-GFP), their differentiated TGCs (GC-R26 and GC-GFP), MEFs and 3T3 cell lines were used. *Psg* specific primers were designed as per manufacturers instructions, spanning a region 300 bp in the TSS upstream regions of *Psg22* and *Psg23*. *Psg22* primers were located at -151 bp from the Transcriptional Start site, and the *Psg23* primers were located -125 bp from the Transcriptional Start site. Primers used are described (Table 2.1.). Cultured cells were exposed to *in situ* chromatin nuclease digestion, genomic DNA was purified and qRT-PCR analysis was performed as per protocol. Three biological replicates and three technical replicates for each cell line was evaluated using the online EpiQ Chromatin Kit Data Analysis Tool, provided with the kit. A lower than 65% accessibility result, deems the conformation of the region inaccessible and thus moderately silenced. A 65% or above accessibility result, deems the chromatin state accessible and active, meaning this region is minimally silenced, or not silenced at all. The percentage of chromatin accessibility for the *Psg22* 300 bp TSS upstream region, in a variety of cell lines are shown (Fig:3.24.A).

A



B



Accessibility	Chromatin Structure	Potential of Epigenetic Silencing
95-100%	Fully accessible	Not silenced
65-95%	Mostly accessible	Low level of silencing
20-65%	Low accessibility	Moderately silenced
0-20%	Highly inaccessible	Completely silenced

Figure 3.24: Quantification of chromatin accessibility in *Psg22* and *Psg23* promoter regions in TSCs (TS-GFP and TS-R26), differentiated TGCs (GC-GFP and GC-R26), MEFs and 3T3 Cells. Biorad EpiQ Chromatin Accessibility Assay was used to determine percentage accessibility of chromatin using *Psg* specific promoter primers with 300 bp amplicons. Unlike the *Psg23* promoter, the *Psg22* promoter region is in the active chromatin conformation in differentiated TGCs. (n=3).

Fibroblast derived MEF and 3T3 cell lines were used as controls, as there is no *Psg22* or *Psg23* expression found in these cell lines and the chromatin in these upstream regions should not be in an active conformation in these cell lines. The chromatin accessibility of the *Psg22* upstream region is found to be 49% in MEFs, and 55% in 3T3 cells, and is in the inactive conformation which was expected. Chromatin accessibility was found to be 36% in the TS-R26 cell line and 54% in the TS-GFP cell line. The conformation in both these TSC lines is the inactive state, which is consistent with low *Psg* expression levels in these cell lines. Analysis of differentiated TGC chromatin conformation has revealed that the *Psg22* 300 bp TSS upstream region is in the active conformation in these cells. The *Psg22* 300 bp TSS upstream region had a 67% accessible chromatin structure in GC-R26 cells, and a 71% chromatin accessible chromatin structure in GC-GFP cells. This demonstrates that the *Psg22* promoter is mostly accessiblenin these TGC populations which is unsurprising as these cells are the primary source of *Psg22* expression, and an open chromatin conformation is expected due to the high levels of *Psg22* transcription.

As a *Psg* 300 bp TSS upstream region control, the *Psg23* 300 bp TSS upstream region chromatin conformation was also investigated. As per the *Psg22* region, the chromatin accessibility of the *Psg23* 300 bp TSS upstream region was found to be inaccessible in MEFs, 3T3 cells, and in both TS cell lines (Fig:3.24.B). *Psg23* chromatin accessibility was 30% in MEF cells, 35% in 3T3 cells, 31% in TS-R26 cells, and 39% in TS-GFP cells. All of these cell lines demonstrated that the *Psg23* promoter was poorly accessible due to an inactive chromatin conformation. In contrast to the *Psg22* promoter region, the *Psg23* promoter chromatin conformation was found to be in the inactive state in differentiated TGCs, giving an accessibility result of 53% in GC-R26 cells, and 57% in GC-GFP cells, just short of the 65% cutoff percentage. These results indicate that the *Psg23* 300 bp TSS upstream region does not undergo a chromatin conformational change as a consequence of TSC differentiating into TGCs, as occurs with the *Psg22* 300 bp TSS upstream region. There is only a slight difference in promoter activity between these two regions in the LacZ-reporter assay, which is surprising given the stark differences in chromatin conformation demonstrated in

TGCs for these regions (Fig:3.24.C). I have found, using the Chromatin Accessibility assay, that the murine *Psg22* upstream region undergoes substantial changes in chromatin accessibility upon TGC differentiation that is not seen in the corresponding region of *Psg23*.

3.2.7 Identification of *Psg22* antisense transcript

I found only slight differences in promoter activity between the *Psg20, Psg22* and *Psg23* regulatory regions which suggests that an alternative mechanism is responsible for the increased *Psg22* expression in TGC. Transcription factor binding analysis of murine *Psg* upstream regions did not suggest an explanation. To address this issue, I investigated a putative enhancer element located upstream of *Psg22* that is not present in the rest of the murine *Psg* family which may be responsible for these high levels of *Psg22* expression. From an extensive search using the available genome browsers, I located an Expressed Sequence Tag (EST) located approximately 5738 bp upstream of the *Psg22* Transcriptional Start Site. This EST was found using the EST track on the UCSC genome browser. Annotated as BY564540, this EST is 417 bp in length and is located on the negative strand. A screen capture from the UCSC browser (Fig:3.25.A), illustrates the location of BY564540 in relation to *Psg22*. Using the BY564540 sequence and the online BLAST programme, I examined whether there were any other regions in the murine *Psg* family locus that possessed a similar EST or region of similarity. The BLAST results indicated that there are three other regions within the *Psg* locus, that had very similar sequence to the BY564540 EST. I named these regions BLAST 1-3. The closest match to the BY564540 sequence was located approximately 9264 bp downstream of *Psg22* stop codon (TGA). This sequence (BLAST 1) was 90% similar to the original EST sequence. Two other BLAST hits, (BLAST 2 and BLAST 3) revealed sequences that were 86% and 81% similar respectively. BLAST 2 sequence was located upstream of *Psg19* and the third BLAST hit, BLAST 3 was located upstream of *Psg25*. Locations of these three BLAST hits, along with the original EST BY564540, and their relative orientations on the mouse *Psg* locus are shown (Fig:3.25.B). All BLAST sequences were found to be in the opposite orientation to the *Psg* genes that they

reside next to. It was found that the BLAST 1 hit, downstream of *Psg22* is in the same orientation as BY564540, whereas the other two BLAST hits, BLAST 2 and 3 are in the opposite orientation. I investigated if BY564540 and transcripts arising from the regions identified by the BLAST analysis are expressed in trophoblast lineages. Using RT-PCR, BY564540 specific primers were designed (Table 2.4.), and the expression of this EST was determined. This BY564540 EST is expressed in trophoblast lineages, but more interestingly, is expressed in a TGC specific manner (Fig:3.25.C). Two TSC lines were tested, and expression of BY564540 was not found in these TSC lines, but when these TSC undergo differentiation, BY564540 EST expression is detected. I have shown its expression in two differentiated TGC lines, EPC tissue, and three stages of placental development (E13, E15, and E17). No expression of this EST can be found in ES cells. I investigated the expression of the three BLAST result regions in trophoblast lineages. BLAST 1-3 region specific primers were designed using the Primer-Blast programme and I examined their expression in the TSC and differentiated TGC. As with the original BY564540 EST, these sequences were also expressed in a TGC specific manner, with no expression found in undifferentiated TSC. RT-PCR products were cloned and sequence verified (GATC, Germany). All RT-PCR amplicon sequences returned were BLASTed against the mouse genome, and sequences corresponded to the exact sequences predicted by the BY564540 EST BLAST results (Fig:3.25.D). The BY564540 EST is expressed, as are the three BLAST regions that are similar to this EST. The 417 bp sequence that is present on the UCSC genome browser was analysed using the online ORF finder software programme, (http://www.ncbi.nlm.nih.gov/projects/gorf/gorf.html), and no coding ORFs were found in any of the three frames tested. This result establishes that this BY564540 EST, is expressed in a TGC-specific manner, has no protein-coding potential, and that this *BY564540* antisense transcript is a noncoding RNA transcript. The fact that *Psg22* is flanked by these two antisense transcripts, may have a role in the upregulation of *Psg22* expression in TGCs in the first half of pregnancy. The next step was to map these non-coding RNA antisense transcripts and to test if these antisense transcripts have a regulatory function in TGC.

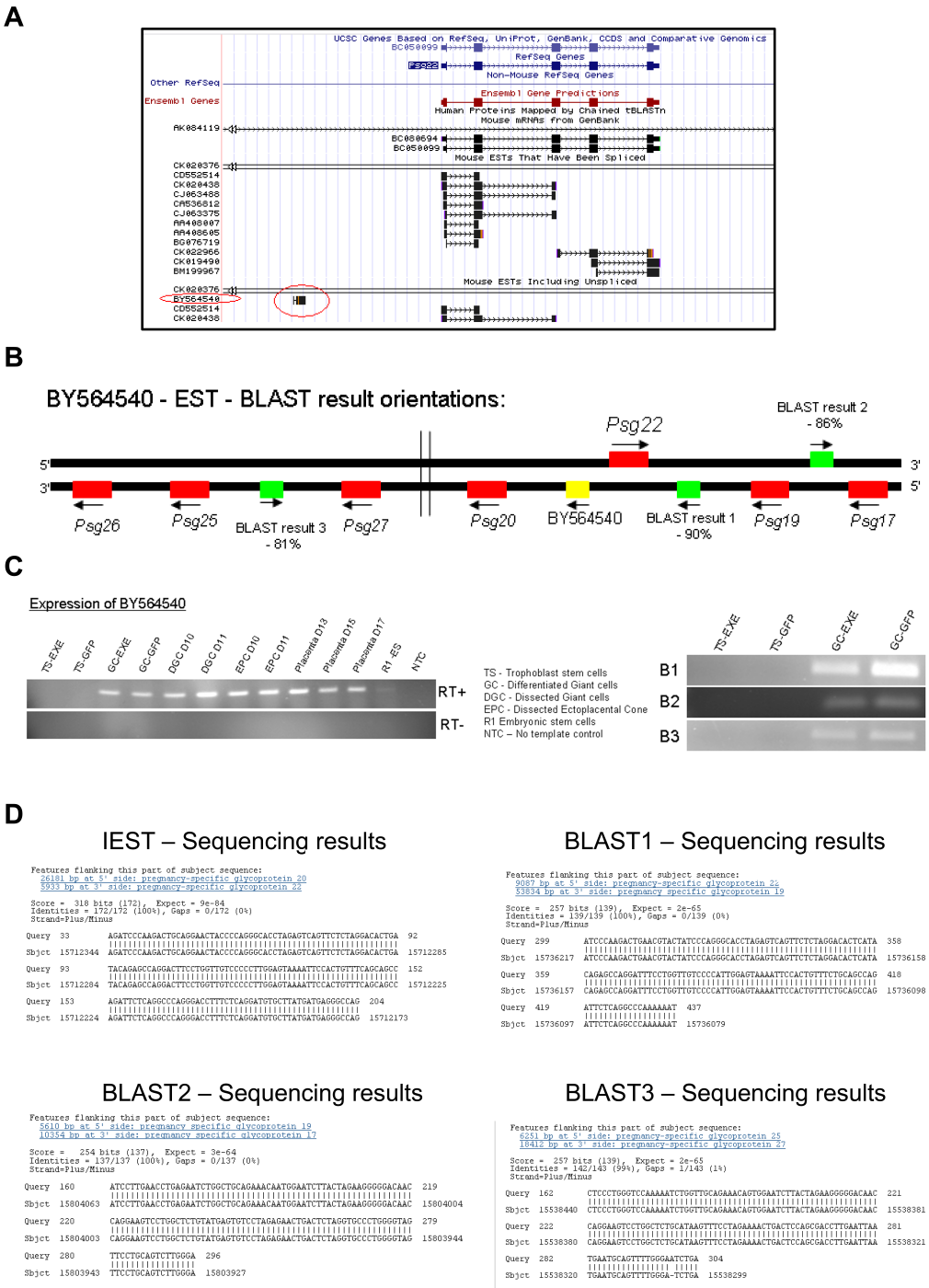


Figure 3.25: Identification of BY564540 EST transcript. (A) BY564540 EST located upstream of *Psg22* exon 1 on UCSC genome browser. (B) BLAST results of BY564540 EST. Three regions highlighted by green boxes contain homologous sequences to BY564540 EST (yellow box). Named BLAST result 1-3, the relative orientations of these BLAST results are shown. (C) BY564540 EST transcript is expressed in TGC but not in TSC. The expression of BLAST 1- 3 (B1-B3) regions is also found in TGC but not in TSC. (D) RT-PCR amplicons were cloned and sequenced, all four transcripts are expressed.

3.2.8 Mapping of the *BY564540* and *BLAST 1* antisense transcripts using primer walking

Following on from the detection of expression of this EST and similar regions, it was necessary to map the structure of these antisense transcripts. I employed a primer walking approach to map the *BY564540* transcript, as an initial attempt at 5'RACE (Rapid Amplification of cDNA Ends), to map the 5' end of this transcript was unsuccessful (data not shown). Antisense transcript specific primers were designed, and used to RT-PCR this transcript in an overlapping manner, to find the transcribed boundaries. Specific RT-PCR primers used for the Primer walking of Antisense transcripts are shown (Tables 2.5. & 2.6.). Transcript specific primer walking primer locations, and the locations and distances of the *BY564540* EST and its similar *BLAST 1* antisense transcript in relation to the *Psg22* locus is illustrated (Fig:3.26.). Using the primer walking method I mapped the *BY564540* and *BLAST 1* antisense transcripts. Amplifying E10 dissected TGC cDNA using specified RT-PCR primer combinations gave an approximate size of *BY564540* antisense transcript as 6148 bp and the approximate size of *BLAST 1* antisense transcript is 6370 bp. Using RT-PCR primer walking, I was able to detect transcription of the *BY564540* antisense transcript to the TSS of *Psg22*. The length and position of the 5' end of this transcript implicates this *BY564540* antisense transcript, as a divergent or bidirectional lncRNA, as its transcription initiates within 1000 bp of the TSS of *Psg22*, the fact that this transcript is over 6 kb in length, and contains no ORFs in any of the three frames analysed. The presence of the second lncRNA antisense transcript, *BLAST 1* lncRNA antisense transcript, downstream of *Psg22*, may have occurred as a result of a duplication event of the *BY564540* lncRNA, when the *Psg22* gene locus was subjected to the inversion event (Fig:3.1.). I hypothesise that the expression of these lncRNA transcripts function in maintaining an open local chromatin conformation, resulting in ease of access of the *Psg22* transcriptional machinery to *Psg22* promoter regulatory regions.

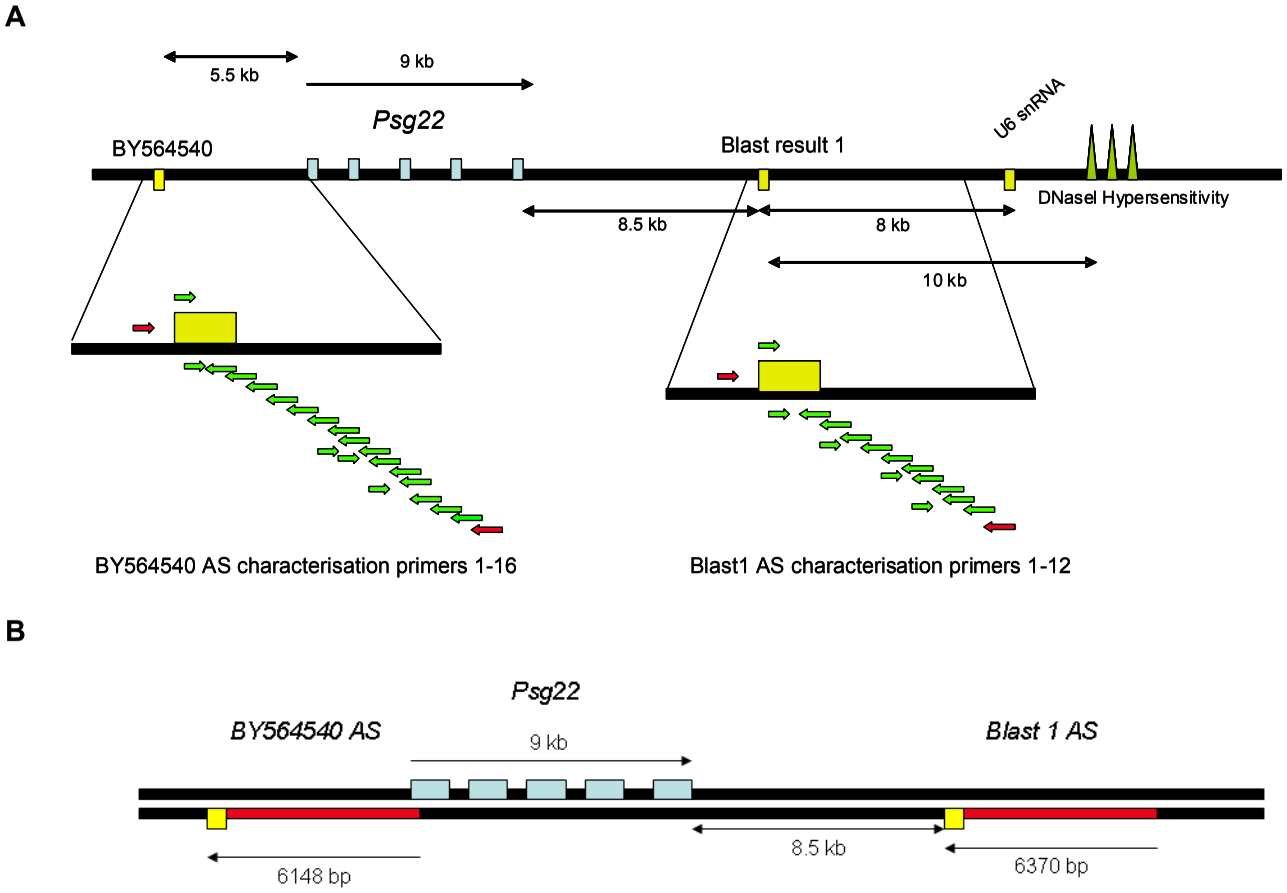


Figure 3.26: Mapping of *BY564540* antisense transcription borders. (A) Schematic of *Psg22* locus, with locations of BY564540 EST and Blast 1 region. Also shown are primer walking oligonuceotide locations. Green arrows indicate positive amplicons and red arrows indicate no amplification. (B) Schematic of *Psg22* locus with approximate sizes of *BY564540* antisense transcript (6148 bp) and *BLAST 1* antisense transcript (6370 bp). Yellow boxes indicate original EST sequences, and red boxes indicate mapped transcripts.

3.2.9 Investigation of *BY564540* antisense transcript expression relative to *Psg22* expression

As the primer walking experiment has demonstrated, the *BY564540* EST antisense transcript is approximately 6 kb in length, and is expressed in differentiated TGCs, as is *Psg22*, suggesting that there is a possibility of this antisense transcript being involved in the regulation of *Psg22* expression. I investigated the relative abundance of the *BY564540* lncRNA antisense transcript relative to the expression of the primary *Psg22* transcript and expression patterns in trophoblast cells and tissues. Employing the same technique used to quantify the expression of the *Psg22* splice variants as described in materials and methods, a specific region of each transcript was dual cloned into a single construct, which was used to construct a standard curve for qRT-PCR analysis. Once both inserts had been correctly cloned and sequence verified, a standard curve was constructed using serial dilutions of the template plasmid as described [251, 252]. Primers used in cloning of the dual transcript vector, and qRT-PCR primers used are described (Table 2.10.). A variety of trophoblast derived cell lines and tissues were used as templates for the qRT-PCR reactions, including two TSC lines (TS-EXE and TS-GFP), their differentiated TGCs, dissected TGCs (E10 and E11), dissected EPC (E10 and E11), and three embryonic stages of placenta (E13, E15 and E17).

The relative quantification of *Psg22* and *BY564540* antisense transcripts can be seen (Fig:3.27.). These results show that this antisense transcript follows the same expression patterns that were found for both *Psg22* transcripts, having higher expression in earlier embryonic time points and expression levels lowering as embryonic development progresses (Fig:3.27.A&B). Reproducing my previous results, the expression pattern of the *Psg22* transcript is the same as in (Fig:3.13.C & Fig:3.14.C). This data shows that there are low levels of expression of both the *Psg22* transcript and the *BY564540* antisense transcript in both TSC lines. This is due to a mixed population of cell types found therein, which was shown by expression of *Eomes, Pl2* and *Tpbpa*. *Psg22* transcript expression increases upon differentiation to TGCs, as does the *BY564540* antisense transcript. The increase of expression of the *BY564540* transcript

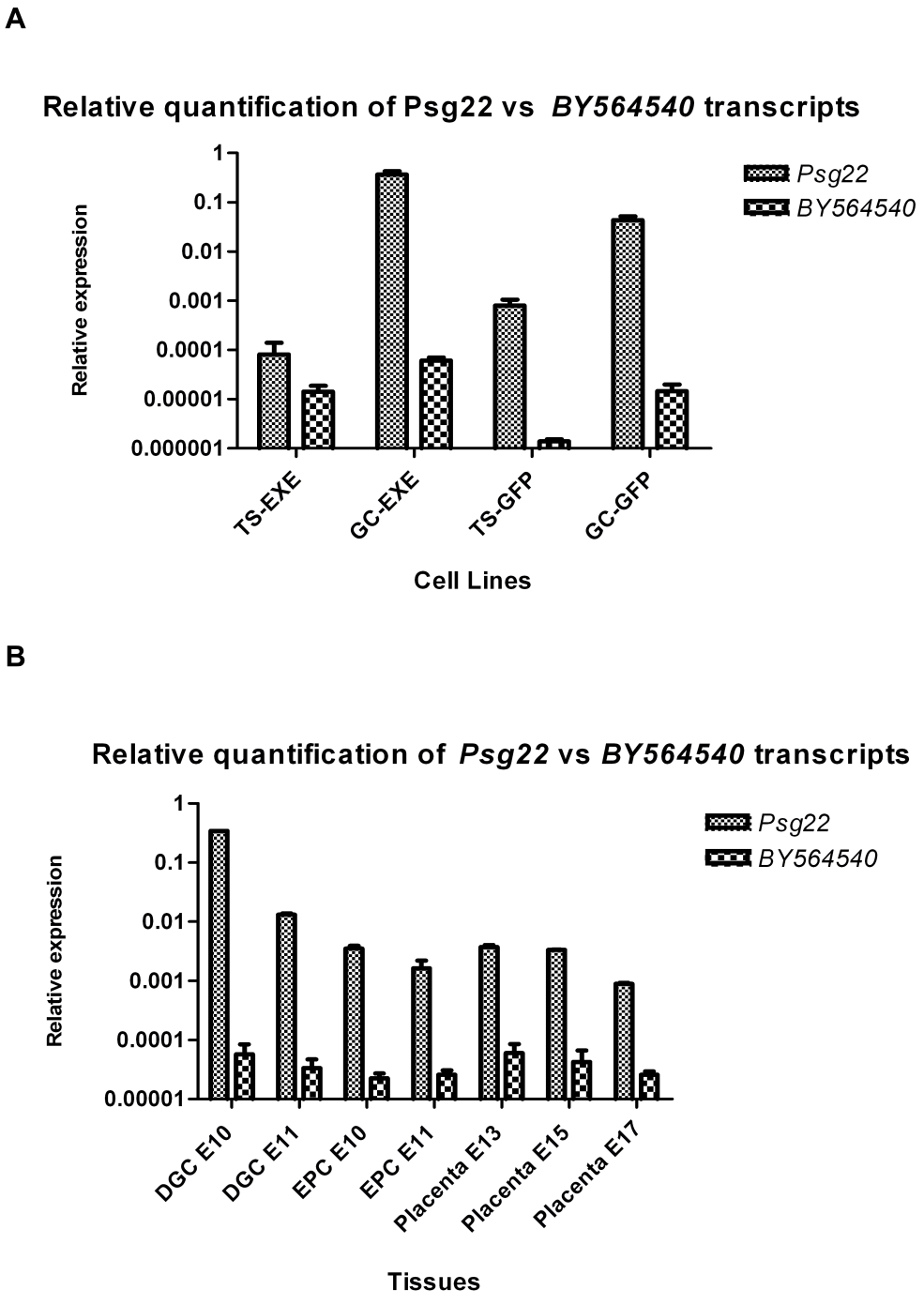


Figure 3.27: Relative quantification of *Psg22* compared to *BY564540* antisense transcript expression. (A) Relative quantification in TSC and differentiated TGC. (B) Relative quantification in trophoblastic tissues. *Psg22* expression is remarkably higher than the antisense transcript in all cell lines tested, with the closest expression found in TSC. (n=3)

in differentiated TGCs, and the mirroring of *Psg22* expression patterns suggests that these two transcripts are co-expressed and this *BY564540* antisense transcript may be regulating the expression of *Psg22* or *vice versa*, in a tissue specific manner. These similar expression patterns are also observed in the trophoblastic tissues tested. The differences in expression levels are quite stark in these tissues, with a nearly 1000 fold difference in expression levels between the two transcripts in E10 dissected TGCs. Expression of *BY564540* decreases from E10 to E11 in dissected TGCs, and in the three embryonic stages of placenta, decreasing as the placenta develops from E13 to E17.

These results demonstrate that the *BY564540* antisense transcript is expressed at low levels in trophoblastic cell lines and tissues, with the highest levels of expression found in dissected TGCs and E13 placental samples, demonstrating a concordant expression pattern with its neighbouring gene, *Psg22*. These concordant expression patterns are further evidence that these lncRNAs may be responsible for the upregulation of *Psg22* expression through an possible epigenetic mechanism of transcriptional regulation. This low level of expression could possibly maintain an open chromatin structure surrounding the *Psg22* locus, which in turn may facilitate in the increased expression of *Psg22*, as the transcriptional machinery involved in *Psg22* expression encounters an open chromatin conformation, and the *Psg22* promoter is easily accessible for the initiation of transcription. The opposite is also possible whereby, the expression of *Psg22* may modulate local chromatin conformatin and regulate the expression of these lncRNAs. Whether these lncRNA transcripts are involved in modulation of local chromatin, is addressed in the next section.

3.2.10 Investigation of chromatin structure and accessibility in *BY564540* and BLAST 1 antisense transcript regions

Continuing the investigation concerning the chromatin conformation associated with the *Psg22* upstream region, which supported the hypothesis for the role of the *BY564540* antisense transcript in the regulation of *Psg22* expression in TGCs, I investigated the conformational states of chromatin of the original BY564540 EST and its Blast 1 result regions. The EpiQ chromatin analysis kit was used for

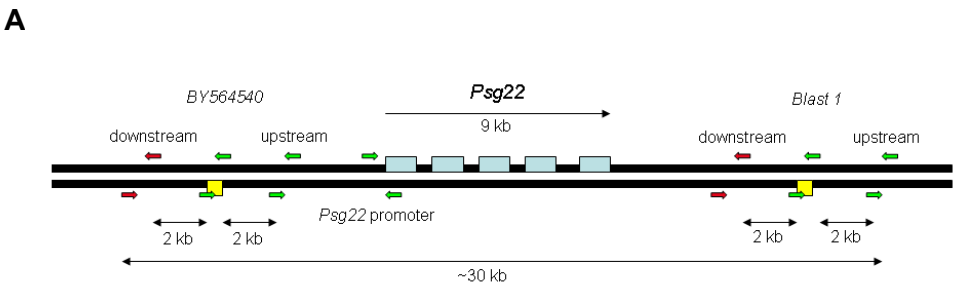
this investigation. Region specific primers were designed as per protocol and are described (Table 2.1.) Primers were designed spanning 300 bp within the original BY564540 EST sequence, also primers were designed in a region 2 kb upstream and 2 kb downstream of the BY564540 EST. Similar primer sets were designed using the BLAST 1 result and surrounding regions as a template for primer design. Primers were designed within the corresponding BLAST 1 result sequence, and also 2 kb upstream and 2 kb downstream of this region. Locations of primers on the *Psg22* locus that were used in this experiment are shown (Fig:3.28.A). These downstream flanking primer sets were used to distinguish between regions that are actively transcribed, in contrast to regions lacking active transcription. The previous primer walking experiment has mapped the regions that are transcribed on both these antisense transcripts (Fig:3.25.).

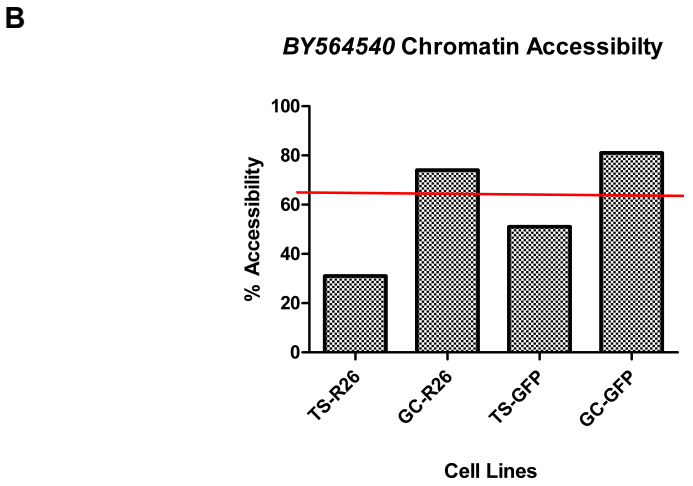
As before, TS-R26 and TS-GFP, differentiated TGC (GC-R26 and GC-GFP) were used as templates. The percentage accessibility of the BY564540 EST region in TSC and differentiated TGCs is shown (Fig:3.28.B). The chromatin conformation of this region was found to be in an inactive state in both TSC lines tested, returning chromatin accessibility of 31% in TS-R26 cells, and 51% in TS-GFP cells. This result is consistent with the extremely low levels of expression of *BY564540* found in TSC lines, which would therefore have an inactive chromatin conformation as a result of this low expression. Congruous with the results obtained from the *Psg22* promoter investigation in TGCs, the region of chromatin associated with the BY564540 EST antisense transcript was shown to be in an open, active conformation, upon differentiation to a TGC fate. Chromatin accessibility in this region was found to be 74% in GC-R26 cells, and 81% in GC-GFP cells. The open conformation of chromatin in this region in TGCs may be due to the expression to the of these antisense transcripts, or conversely, the expression of these antisense transcripts may facilitate in opening the local chromatin conformation.

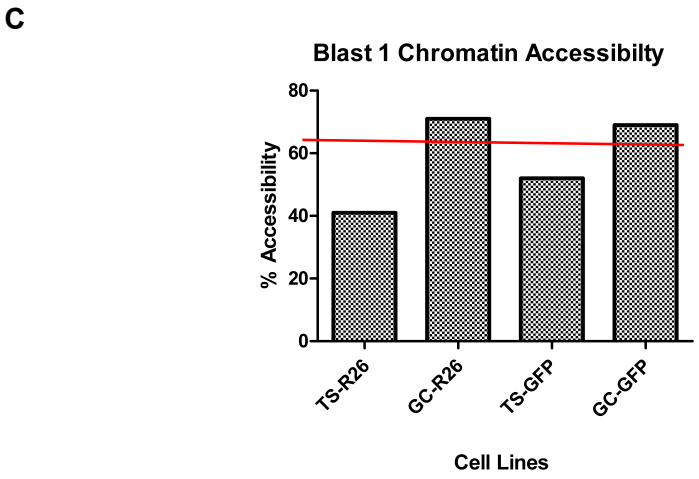
Similar results found in the conformation of chromatin within the BLAST 1 antisense transcript are shown (Fig:3.28.C). As with the *BY564540* antisense transcript, this BLAST 1 region has a closed chromatin conformation in TSC, with a chromatin

accessibility of 41% in TS-R26 cells, and 52% in TS-GFP cells. As before, we can see a conformational change of chromatin, once TSC undergo differentiated into TGC. Chromatin accessibility of 71% in GC-R26 cells and 69% in GC-GFP cells was observed for this BLAST 1 antisense transcript, mirroring the results obtained with the *BY564540* EST antisense transcript. These two antisense transcripts, have demonstrated concordant expression patterns, and are associated with TGC-specific open chromatin states. These data support the hypothesis of the divergent/bidirectional *BY564540* antisense transcript and its related and *BLAST 1* antisense transcript playing a pivotal role in the upregulation of *Psg22* expression or are correlated with it.

To confirm that this active local chromatin conformation is correlated with the expression of these antisense transcripts, I performed the chromatin accessibility assay using primer sets located 2 kb upstream and 2 kb downstream of the *BY564540* and *BLAST 1* antisense transcripts. The upstream primer sets for both antisense transcripts are located within these transcripts and are located in regions where transcription is active. Using the upstream *BY564540* primers, two TSC lines demonstrated a closed chromatin conformation, having chromatin accessibility of 40% in TS-R26 cells, 51% in TS-GFP cells. I found that there is a TGC-specific opening of chromatin conformation in the region 2 kb upstream of the original BY564540 EST (Fig:3.29.A). A chromatin accessibility of 72% was found in GC-R26 cells and the GC-GFP cell line demonstrated a chromatin accessibility of 83%. Similar results were found in the 2 kb upstream region of the *BLAST 1* antisense transcript (Fig:3.29.B), as there is the same chromatin conformational change in this region upon TGC differentiation. In TSC, there is a chromatin accessibility of 43% (TS-R26) and 50% (TS-GFP) respectively, whereas the local chromatin opens considerably in TGC, with chromatin accessibility of 77% in GC-R26 and 74% in GC-GFP cell lines in this upstream *BLAST 1* region. Taking into account that these upstream regions are located within the *BY564540* and *BLAST 1* transcripts, it is not surprising that there is a similar chromatin conformation within these upstream regions. These results demonstrate that regions which are actively transcribed and contain these antisense transcripts are associated with an open conformation of the surrounding chromatin, rather than it being a feature of







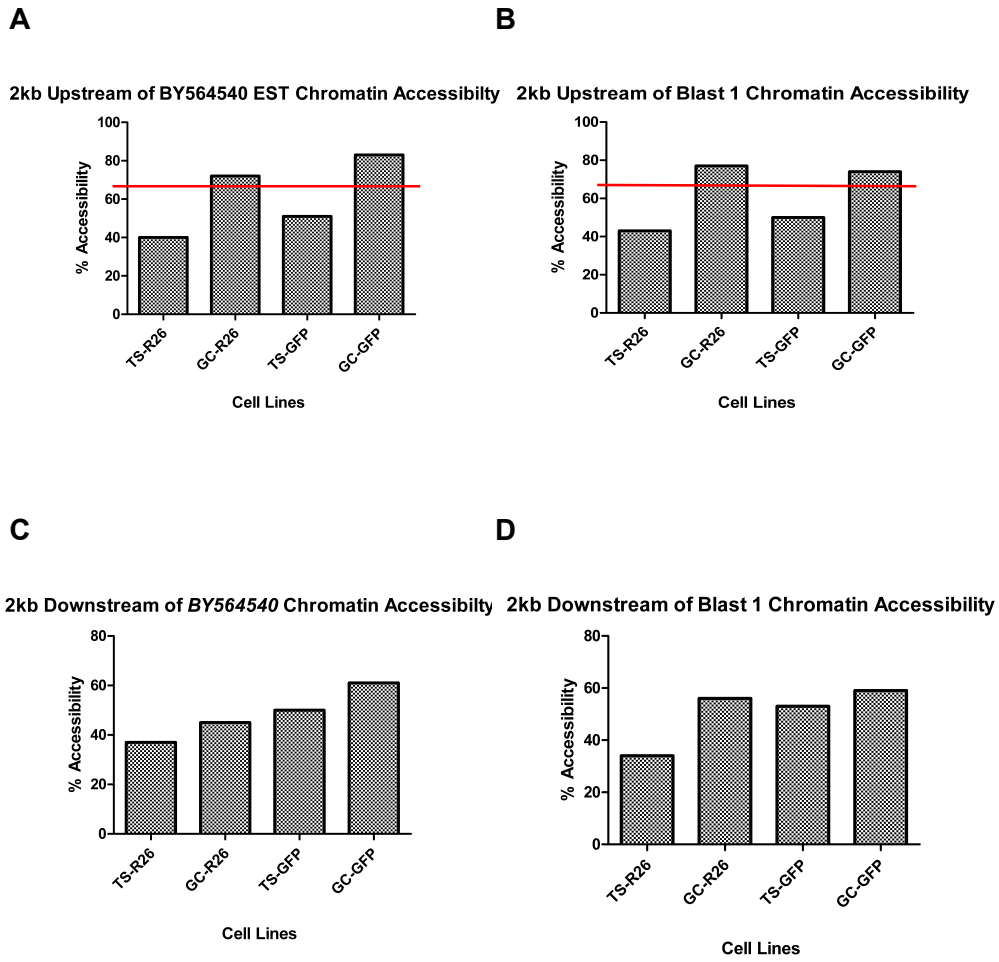
Accessibility	Chromatin Structure	Potential of Epigenetic Silencing
95-100%	Fully accessible	Not silenced
65-95%	Mostly accessible	Low level of silencing
20-65%	Low accessibility	Moderately silenced
0-20%	Highly inaccessible	Completely silenced

Figure 3.28: Quantification of chromatin accessibility in *BY564540* and *BLAST 1* antisense transcripts in TSC, and differentiated TGC. (A) Locations of primers used in chromatin accessibility assay on *Psg22* locus. (B) Quantification of chromatin accessibility in *BY564540* antisense transcript. (C) Quantification of chromatin accessibility in *BLAST 1* antisense transcript. Biorad EpiQ Chromatin Accessibility Assay was used to determine percentage accessibility of chromatin using antisense specific specific promoter primers with 300 bp amplicons. (n=3)

local chromatin. These results confirm that the chromatin accessibility in these regions is correlated with the expression of two novel antisense transcripts, in a TGC-specific dependent manner.

I investigated whether chromatin conformational state in these regions is dependent on expression of these antisense transcripts and is not just a local chromatin feature that spans several kb. I investigated, whether there is chromatin conformational change in TGC, in a region that is not associated with transcription. I used regions 2 kb downstream of these antisense transcripts and designed region specific primers as per protocol. The conformation of the chromatin in a region 2 kb downstream of *BY564540* antisense transcript that does not have an associated transcript is shown (Fig:3.29.C). Both TSC lines have an inactive or closed chromatin conformation, with chromatin accessibility at 37% (TS-R26) and 50% (TS-GFP). Upon differentiation to TGC, we see a slight increase in the chromatin accessibility, 45% (GC-R26) and 61% (GC-GFP), but this increase in accessibility, is not enough to deem the chromatin in an accessible state (below 65%). The same pattern was found in the region 2 kb downstream of the *BLAST 1* antisense transcript, that upon differentiation to a TGC fate, there is no change to the overall conformation of chromatin in this region. The chromatin accessibility observed in both TSC lines was 34% (TS-R26) and 53% (TS-GFP), while the accessibility of chromatin of 56% (GC-R26) and 59% (GC-GFP) was observed in both TGC lines for this downstream region (Fig:3.29.D). This demonstrates that regions which are not actively transcribed are not associated with an open chromatin conformation.

This difference between TSC and TGC chromatin conformation in these regions is correlated to low level expression of the *BY564540* and *Blast1* lncRNA antisense transcripts, and is associated an active open chromatin conformation in the *Psg22* promoter region. This association is not present in the *Psg23* promoter, which demonstrates that this is *Psg22*-specific rather than promoter associated chromatin opening. This maintenance of an active chromatin state in the promoter region of *Psg22* by low level expression of the *BY564540* antisense transcript, further supports the hypothesis that the *BY564540* antisense transcript is a bidirectional lncRNA, with



Accessibility	Chromatin Structure	Potential of Epigenetic Silencing
95-100%	Fully accessible	Not silenced
65-95%	Mostly accessible	Low level of silencing
20-65%	Low accessibility	Moderately silenced
0-20%	Highly inaccessible	Completely silenced

Figure 3.29: Quantification of chromatin accessibility in a region 2 kb upstream and 2 kb downstream of *BY564540* and *BLAST 1* antisense transcripts in TSC, and differentiated TGC. Quantification of chromatin accessibility in (A&B) 2 kb upstream of *BY564540* and *BLAST 1* antisense transcripts; and (C&D) 2 kb downstream of *BY564540* and *BLAST 1* antisense transcripts. Biorad EpiQ Chromatin Accessibility Assay was used to determine percentage accessibility of chromatin using specific primers with 300 bp amplicons. (n=3)

a role in the enhancement of *Psg22* expression in a cell-specific manner. The exact mechanism of how this enhancer RNA (eRNA) functions is yet to be determined, and future work is needed to elucide the exact mechanisms. I hypothesise a mechanism that is similar to the the enhancer mechanism proposed by Rinn *et al*, [221], (Fig:1.9.D), in which chromosome looping of these antisense transcripts, maintains an active local chromatin state, enabling the *Psg22* transcriptional machinery access to the *Psg22* promoter. This epigenetic transcriptional regulation of *Psg22* is a novel mechanism that has to date not been described in the PSG or CEACAM families.

Chapter 4

Discussion and future directions

4.1 A review of human and rodent PSG loci

The multigene *PSG* family, is a rapidly evolving subset of placenta-specific hormones that has been shown to be undergoing positive selection [111]. To fully understand the expansion and evolution of this family, correct gene sequences and their locations at the *Psg* locus are needed. Using new mouse genome assemblies available on the publically accessible genome databases, I compiled an up-to-date accession table of all known human, rat and murine PSGs. Using the correctly annotated murine *Psg* sequences, I was able to discern the genomic length, exon structure, gene orientation, TSS, CDS and locus coordinates for all murine *Psg* genes. I determined ORF length and domain structure of each corresponding Psg protein. From these data, I produced an updated map of the previously predicted *Psg* locus [146]. The discovery of a recent gene inversion event of *Psg22* within the *Psg* locus is interesting as it may explain the high levels of *Psg22* expression relative to the murine *Psg* multigene family and provides new information concerning the evolution of the murine *Psg* genomic locus structure and organisation. It is unknown when this inversion event occured but it is common to at least two murine strains. These correct gene loci maps and accession table have produced a detailed description of the entire rodent *PSG* family and will aid in further studies of PSG expression and function.

Within both mouse and rat *PSG* families, there is a cluster of members that are flanked by the *Mill1* and *Mill2* genes. In the murine *Psg* family, 11 of 17 *Psgs* are located within this cluster. In contrast, only two of eight rat *PSGs* are located within this region. This suggests that there was either an expansion of the murine *Psg* family within the major *Psg* cluster relative to the rat *PSG* family, or that the rat *PSG* family experienced a contraction within this cluster. It is hypothesisd that expansion of the murine *Psg* gene family suggests that this multigene family is under selection both for increased gene dosage and diversification of function [146]. To gain a better insight into the evolution of rodent *PSGs*, a neighbour-joined pairwise-comparison phylogenetic tree of murine and rat *PSG* CDS were constructed. Co-branching of certain murine and rat *PSGs*, upstream of the *Mill1/2* flanked rodent *PSG* cluster, in the multi-species phylogenetic analysis suggests that these regions are syntenic, and that there are orthologous relationships between members of these species. The identification of 5 rodent *PSG* orthologous relationships that are common to this region in both species is important for the reliable prediction/extrapolation of gene function. To date there has been no human *PSG* orthologues found. The orthologous relationship between *PSG36* and *Psg24* is also supported as both contain five N domains [146].

I employed a *Psg* specific probe and southern hybridisation, to screen a mouse 129/Sv PAC library and obtained a number of *Psg* containing PAC clones. The PAC3 clone (647-D4) contains a region of the *Psg* locus (*Psg26 - Psg22*), which was confirmed by PCR characterisation. End sequencing of the PAC3 clone, has revealed that the *Psg22* inversion event is also common to the 129/Sv mouse strain, and is not a strain specific evolutionary event. This PAC3 clone was used to clone the 2 kb regulatory regions of *Psg20, Psg22,* and *Psg23*, that were used in LacZ-reporter assays in this thesis. These *Psg* containing clones can be used in future research as sources of isogenic homology arms used to construct individual *Psg* KO vectors or a locus KO vector.

4.2 PSG expression profiles and non-plancental PSG expression

To facilitate the investigation of *Psg* expression in trophoblast lineages, an *in vitro* cell culture model that expresses endogenous *Psg* was needed. I employed a 6 day FCM withdrawal method to differentiate two TSC lines (TS-EXE and TS-GFP) into predominantly TGC populations and surveyed *Psg22* expression in both undifferentiated and differentiated states. I found that there was a clear upregulation of *Psg22* expression upon differentiation towards a TGC fate. This *Psg22* expression was comparable to the expression levels found in E15 placenta. It has been previously shown that RA induces differentiation towards a TGC fate [101], although RA treated TSCs failed to produce a high level of endogenous *Psg22* expression in comparison to the 6 day FCM withdrawal protocol, but may be used in conjunction with FCM withdrawal, to enhance TGC-specfic differentiation. This cell culture model of endogenous *Psg22* expression with expression levels comparable to that of placental tissue can be used to further elucidate the expression, regulation and functions of *Psgs in vitro*. This model can be utilised to determine the exact role of *Psgs* in trophoblast development and in TGC differentiation.

To support previous data concerning murine *Psgs* expression in the placenta, I have shown that two previously uncharacterised *Psgs*, (*Psg31* and *Psg32*), were expressed in E15 placental tissues. Using overlapping primers, and sequence analysis of cloned RT-PCR amplicons, I was able to map both *Psg31* and *Psg32* transcripts. These cloned sequences were aligned against the *Psg* locus and I determined the correct exon and domain structure of both these genes. I found that *Psg31* has 10 exons and is composed of an N1-N1*-N2-N3-N4-N5-N6-N7-A domain structure, which supports previously predictions that *Psg31* has evolved from a duplication of the entire *Psg30* gene and a subsequent duplication of the N1 domain [148]. Sequencing analysis revealed that *Psg32* contains 5 exons, and has a N1-N2-N3-A domain structure. I have shown that *Psg32*, (previously *Cea6* or *Psg-ps1*) [106], is not a pseudogene, and is expressed in murine placenta. The expression of *Psg31* and *Psg32*

adds these genes to the list of placentally expressed *Psgs*.

Previous studies of *Psg* gene expression in mouse pregnancy indicated that different family members exhibit different expression levels between E11 and E18, suggesting the possibility of divergent functions [146]. Using comprehensive semi-quantitative expression studies I have generated an expression profile of murine *Psgs* in a variety of trophoblast lineages and cell lines. RT-PCR expression surveys revealed that there are a number of *Psgs* expressed in two TSC lines, but that upon differentiation to TGC, *Psg22* is the most abundant transcript. I found similar *Psgs* expressed in E5 blastocysts, with *Psg22* being the most abundant transcript in E11 blastocyst outgrowths, which have high levels of TGCs. This supports data that has shown that *Psg22* is expressed from E5.5 in the developing embryo to the remainder of the gestational period in the murine placenta, with highest levels of expression been found in TGC [49, 165].

These results point to a differentiation-led shift in *Psg* expression between undifferentiated TSC and differentiated TGCs in two cell lines and primary blastocyst cultures. Quantitative expression analysis by qRT-PCR confirmed this high expression of *Psg22* in TGCs relative to expression found in TSCs. The expression of *Psg19*, *Psg21*, *Psg22* and *Psg23* was quantified in TSC, TGC, dissected TGC, dissected EPC, and placental samples. *Psg19* and *Psg22* are closely related, and have shown similar expression patterns in these tissues. *Psg21* and *Psg23* are also closely related and also share similar expression patterns. It was found that *Psg22* has the highest expression levels in TGC and dissected trophoblastic tissues, when compared to *Psg19*, *Psg21*, and *Psg23*. These data support previous studies which have shown that *Psg21* and *Psg23* gene transcripts together constitute the bulk of *Psg* gene expression in the SpT, and that *Psg22* constitutes the majority of *Psg* expression in the first half of pregnancy [191].

Furthermore, these data demonstrate that *Psgs* have the same expression patterns *in vitro* as *in vivo* and *Psg* genes display developmentally regulated tissue-specific and cell-specific expression patterns. The importance of describing individual *Psg* family member expression is also confirmed in the predominant *Psg22* expression

in these TGC populations, leading us to believe that *Psg22* may have a specific individual alternative function in early placental development that differs from the remaining *Psg* genes in TGC and time-points in development. These differences in the level and developmental timing of expression of different mouse *Psgs* implicate a divergence of PSG function, although this cannot be confirmed as only four of 17 murine *Psg* were investigated [148, 191]. In summary, expression levels of *Psg* genes in placental tissues of different developmental stages revealed dramatic differences in the developmental expression profile of individual *Psg* family members. Overall the expression data in this study matches well with previous analyses of the distribution of *Psg* transcripts in placental tissues and exhibit further the important role of *Psg22* in early placental development. This expression data will aid in functional studies of this complex gene family.

Non-placental cell expression of certain *Psgs* was found previously in FAE in the GIT, and in the brain [192, 193]. Non-placental *PSG* expression was confirmed in human and mouse GIT tissues by RT-PCR and qRT-PCR, implying a wider role of *PSG* functionality, than one restricted to the placenta. Various *Psg* transcripts were found to be expressed in the GIT of the mouse from the oral cavity to rectum. These results were supported by qRT-PCR which confirmed murine *Psg* expression in esophagus and ascending colon, although this GIT expression was not as high as placental *Psg* expression. This lower level of expression could be due to the fact that to induce a Th2 response in the placenta, a higher dosage of *Psg* is needed than in the GIT, and may be similar in humans although to date there is no evidence for this. Using RT-PCR cloning screens, human PSG expression was also detected in esophageal tissue. The expression of *PSGs* in the human and murine GIT furthers supports the hypothesis of *PSGs* involvement in oral tolerance, and mucosal immune modulation [192]. The expression of human *PSG* in GIT tissues, and comparative mouse *Psg* GIT expression, suggests that these *PSGs* have a conservation of function in both mice and human GIT. The esophagous is a novel site of *PSG* expression, showing that *PSG* expression is not placenta-specific in mice or humans. The function of *PSG* in the GIT needs to be elucidated, and will give a new direction to *PSG* functional research, concerning

regulation of immune and inflammatory mediators in the GIT which would promote a tolerogenic response to commensal bacteria.

The highly expressed *Psg22* was found to have an alternative splice variant, by RT-PCR. These transcripts share concordant expression patterns and are expressed in TGCs. This truncated alternative splice variant has the N1 domain spliced out. Using qRT-PCR, the relative expression of these splice variants was quantified. It was found that the alternative splice variant is expressed at much lower levels than the primary transcript in trophoblast tissues and cell lines. This expression profile suggests that this truncated *Psg22* variant may have a functional relevance to TGCs due to the upregulation of its expression in these cell lines. With the discovery of this high rate of *Psg22* transcription, the levels of translation of the Psg22 protein needed to be established. Since there is a lack of mAbs that specifically detect endogenous Psg22 protein, the ribosome loading of Psg22 transcripts was investigated, as ribosomal loading of transcripts is a good indicator of protein translation. Utilising sucrose gradients and polysome fractionation techniques [253], it was clearly shown that *Psg22* transcripts were indeed heavily loaded with ribosomes, as the majority of *Psg22* transcripts were found in the fractions containing the Polysome bound mRNAs. This is indicative that these transcripts are translated. It is necessary to generate a specific anti-Psg22 antibody to determine the levels of endogenous Psg22 protein *in vivo*.

4.3 Psg22 induces TGFβ1 in monocytes and macrophages

Protein was generated from the two *Psg22* splice variants to investigate whether deletion of the N1 domain affects Psg22 function, as Psg22N1A has been shown to induce TGFβ1from peritoneal macrophages [165]. It was found that both of these proteins induce the release of TGFβ1 from monocytes and macrophages, and encode for proteins with similar function. Generation of recombinant individual domain mutant proteins would be required to discern which region of the Psg22 protein is responsible for this TGFβ1 upregulation. Previous reports have shown that Psg23N1A, and *Psg19* up-regulate TGFβ1 in these cells [170, 265]. It is not

surprising that *Psg22* induces TGFβ1, as *Psg22* and *Psg19* are very similar proteins (Fig:3.2.A). I have shown that Psg22 treatments have upregulated TGFβ1 at the protein level, although I have not investigated whether there is an upregulation of TGFβ1 at the transcriptional level. These data suggests a role for *Psg22* in angiogenesis and immunomodulation as TGFβ1 is an pro-angiogenic, anti-inflammatory cytokine and follows the hypothesis that *Psgs* function as immunoregulators during pregnancy [12]. The treatment of monocytes/macrophages with recombinant murine Psg22 leads to upregulation of the anti-inflammatory cytokine TGFβ1, which has been implicated in the enhancement of Th2-type immune responses [12]. It is hypothesised that *Psg22* expression in early pregnancy may be important for the development of the trophoblast not only by stimulating maternal immune cells to produce angiogenic growth factors but also by direct effects on endothelial cells to promote vascular expansion and development [165].

To asses this function of Psg22, a knock-down of *Psg22* expression *in vitro* was attempted, using two *Psg22* shRNA constructs. These shRNAs were tested in two TGC lines and have shown to generate knockdown of *Psg22* expression *in vitro*. Following on from this, packaging these *Psg22* shRNA vectors into a lentiviral delivery system, and transfect post-fertilisation embryos or ES cells with these vectors, and implant these transfected embryos/cells into pseudo-pregnant female recipient mice to produce chimeric or transgenic neonates which posses a *Psg22* knockdown *in vivo* [249]. This will enable us to utilise these *Psg22* shRNA constructs in future research to produce a knockdown *Psg22* phenotype *in vivo* and investigate the implications of reduced levels of Psg22 protein on pregnancy outcomes. Due to time and financial constraints it was not possible to test these shRNA vectors *in vivo*. The fact that there are 17 murine *Psgs*, which are very similar to each other, and the high probability that murine *Psg* share a common function, may make a single knockdown of an individual *Psg* undetectable in regards to a loss of function phenotype. A complimentary targeted deletion of the major *Psg* cluster flanked by *Mill1* and *Mill2*, may be needed to obtain a knockdown phenotype that is severe enough and not counteracted by the functions of the remaining untargeted *Psg* members. Due to time

constraints I was unable to pursue this experiment.

4.4 *Psg22* regulatory regions exhibit low levels of promoter activity *in vitro*

To elucidate the mechanisms responsible for the regulation of *Psg22*, a 2 kb region containing the predicted regulatory region was analysed to detect transcription factor binding sites that could explain this high level of expression of just one of 17 mouse *Psgs* in the first half of pregnancy. Using transcription factor binding site analysis software, the frequencies of putative transcription factor binding sites of 15 TGC associated transcription factors on the 17 murine *Psg* 2 kb upstream regions were analysed. It has been previously reported that the minimal promoter region of all human *PSG* genes contains a putative Retinoic Acid Responsive Element (RARE) which has been shown to facilitate RXRα transcriptional activation of *PSG5* promoter [200]. There is conservation of this SP1-RXRα (RARE) site in murine *Psgs*: *Psg17*, *Psg19*, *Psg20* and *Psg26*. The fact that *Psg22* does not possess this canonical regulatory region due to a SNP within this region, implies that this mutation in the *Psg22* promoter region has possibly selected against this RXRα site, which implicates the involvement of a different regulatory mechanism that works independently of the SP1/RXR signalling mechanism that is present in all human *PSG* and four of the murine *Psgs*.

From the transcription factor binding analysis (Table 3.1.), it was found that there are 4 other RXR sites present along this 2 kb region *Psg22* promoter region, and it was also shown that RA treatment does induce *Psg22* expression in TSC, demonstrating that RXR signalling regulates *Psg22* expression (Fig:3.7.). There are putative transcription factor binding sites for RxRα in every murine *Psg* 2 kb regulatory region, indicating that this regulatory mechanism is conserved in the mouse as in the human.

Transcription factor binding site analysis did not reveal specific transcription

factor binding sites that would distinguish the *Psg22* promoter from the 16 other *Psgs* promoters. LacZ-promoter-reporter assays also demonstrated that the *Psg22* promoter possessed promoter activity levels similar to *Psg20, Psg21* and *Psg23* in JAR cells. These results suggest that there is a low level of promoter activity associated with *Psg* promoter regions *in vitro*. Although this low level of promoter activity does not properly address the differences in the individual levels of *Psg* expression. It was hypothesised that a differential regulatory method of inducing *Psg22* expression that enhances basal *Psg22* induction by promoter regions may exist. This alternative mechanism could effect individual *Psg* expression levels as there only a slight difference in LacZ expression induced between the *Psg* regulatory regions tested, although there is a difference in expression levels between these *Psg*, both spatially and temporally. Analysis of the conformation of chromatin surrounding the *Psg22* and *Psg23* promoters led to the discovery that the *Psg22* promoter possess heterochromatin in TGC but not TSC, whereas the *Psg23* promoter had its chromatin in a closed state in both TSC and TGC. This led to the hypothesis that there was an alternative unknown mechanism that is responsible for the upregulation of *Psg22*.

Human *PSG* regulation is not only controlled at the transcriptional level via DNA binding factors, it has been shown that *cis* and *trans* acting negative elements repress *PSG5* transcription, irrespective of the cell type [205]. The same kind of mechanisms could control mouse *Psg* regulation. These findings are consistent with the hypothesis that the differences between TATA-containing and TATA-less promoters might allow them to respond to a different subset of activators and or repressors [269]. It is necessary to investigate the role of *cis*/*trans* acting regulatory sequences, epigenetic modulation in the upregulation of *PSG* genes during trophoblast development [181] to provide a better understanding of the regulation of these genes.

4.5 TGC-specific *BY564540* and *Blast 1* antisense transcript expression is correlated to local open chromatin conformation and high expression levels of *Psg22*

Upon further investigation of the *Psg22* locus, an EST sequence was found upstream of the *Psg22* TSS which is transcribed in an antisense direction. Using bioinformatical, RT-PCR and sequencing approaches, it was found that this *BY564540* EST is expressed in a cell specfic manner with expression detected in TGC but not TSC. Three other regions that are similar to the EST sequence were found on the *Psg* locus. Blast results revealed the sequence with the closest similarity to *BY564540* was found to be downstream of the 3' end of *Psg22*. The length of these two antisense transcripts was discovered to be approximately 6 kb, with transcription of the *BY564540* antisense transcript starting within a few hundred base pairs of the *Psg22* TSS. Neither of these antisense transcript possess an ORF in any three frames analysed, indicating that these antisense transcripts are lncRNAs. The presence of this second antisense transcript, *BLAST 1* lncRNA antisense transcript, downstream of *Psg22*, may have occurred as a result of a duplication event of the *BY564540* lncRNA, when the *Psg22* gene locus was subjected to the inversion event.

Relative quantitative expression analysis revealed that this *BY564540* lncRNA antistranscript is expressed in low levels compared to the *Psg22* transcript in a variety of TGC lineage tissues. This qRT-PCR analysis also revealed that the *BY564540* transcript is expressed in a concordant expression pattern to *Psg22*. The expression of *BY564540* lncRNA antisense transcript is ten fold higher in TGC than in TSC, and it is hypothesised that this TGC-specific antisense transcription is correlated with the chromatin conformational change in this region and to the region surrounding the *Psg22* 2 kb regulatory regions. These results show that upon differentiation of TSC to TGC, this region of chromatin undergoes a conformational change from a closed inactive state into a open accessible state that would facilitate the upregulation in expression of neighbouring genes due to the ease of accessibility of transcriptional machinery within this region. Further chromatin analysis of these regions revealed

that in regions where the *BY564540* and *BLAST1* lncRNA antisense trancripts are transcribed, they are associated with an open or active chromatin conformation, and downstream regions, which were non-transcriptionally active, are in the closed conformation. This demonstrates that expression of *BY564540* and *BLAST1* lncRNA antisense transcripts is dependent on TGC differentiation and is associated with open local chromatin conformation. I hypothesise that the high levels of *Psg22* found in TGC, are correlated with the transcription of these *BY564540* and *BLAST1* lncRNA antisense trancripts and the open conformation of local chromatin in the *Psg22* locus.

Due to the fact that these transcripts are non-coding, show concordant expression patterns with neighbouring genes, and are transcribed in a bidirectional antisense manner, it is concluded that these lncRNA antisense transcripts are enhancer RNAs (eRNAs). The open chromatin conformation that is associated with the expression of these antisense transcripts may facilitate the easy access of regulatory machinery to the *Psg22* promoter, and suggests a novel epigenetic regulatory mechanism that to date has not been described in relation to murine *Psg* transcriptional regulation. The exact mechanism in which these eRNAs function is yet to be determined, and future research is needed to elucidate this.

Chapter 5

Bibliography

Bibliography

[1] Springer MS, Murphy WJ, Eizirik E, O'Brien SJ. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proceedings of the National Academy of Sciences of the United States of America. 2003 Feb;100(3):1056–61. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=298725&tool=pmcentrez&rendertype=abstract.

[2] Rawn SM, Cross JC. The evolution, regulation, and function of placenta-specific genes. Annual review of cell and developmental biology. 2008 Jan;24(June):159–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18616428 http://www.annualreviews.org/doi/abs/10.1146/annurev.cellbio.24.110707.175418.

[3] Georgiades P, Ferguson-Smith AC, Burton GJ. Comparative developmental anatomy of the murine and human definitive placentae. Placenta. 2002 Jan;23(1):3–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11869088.

[4] Aubuchon M, Schulz LC, Schust DJ. Preeclampsia: animal models for a human cure. Proceedings of the National Academy of Sciences of the United States of America. 2011 Jan;108(4):1197–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3029729&tool=pmcentrez&rendertype=abstract.

[5] Hemberger M, Cross JC. Genes governing placental development. Trends in endocrinology and metabolism. 2001;12(4):162–8. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/11295572.

[6] Carter aM. Animal models of human placentation–a review. Placenta. 2007 Apr;28 Suppl A:S41–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17196252.

[7] Watson ED, Cross JC. Development of structures and transport functions in the mouse placenta. Physiology (Bethesda, Md). 2005 Jul;20:180–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15888575.

[8] El-Hashash AHK, Warburton D, Kimber SJ. Genes and signals regulating murine trophoblast cell development. Mechanisms of development. 2009;127(1-2):1–20. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2865247&tool=pmcentrez&rendertype=abstract.

[9] Rossant J, Cross JC. Placental development: lessons from mouse mutants. Nature reviews. 2001 Jul;2(7):538–48. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11433360.

[10] Soares MJ. The prolactin and growth hormone families: pregnancy-specific hormones/cytokines at the maternal-fetal interface. Reproductive biology and endocrinology. 2004 Jul;2:51. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=471570&tool=pmcentrez&rendertype=abstract.

[11] Trowsdale J, Betz AG. Mother's little helpers: mechanisms of maternal-fetal tolerance. Nature immunology. 2006 Mar;7(3):241–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16482172.

[12] Motrán CC, Diaz FL, Montes CL, Bocco JL, Gruppi A. In vivo expression of recombinant pregnancy-specific glycoprotein 1a induces alternative activation of monocytes and enhances Th2-type immune response. European journal of immunology. 2003 Nov;33(11):3007–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14579269.

[13] Ha CT, Wu J, Irmak S, Lisboa FA, Dizon AM, Warren J, et al. Human pregnancy specific beta-1-glycoprotein 1 (PSG1) has a potential role in placental vascular morphogenesis. Biology of reproduction. 2010 Jul;83(1):27–35. Available from: http://www.biolreprod.org/content/83/1/27.short http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2888962&tool=pmcentrez&rendertype=abstract.

[14] Chaouat G, Ledée-Bataille N, Dubanchet S, Zourbas S, Sandra O, Martal J. TH1/TH2 paradigm in pregnancy: paradigm lost? Cytokines in pregnancy/early abortion: reexamining the TH1/TH2 paradigm. International archives of allergy and immunology. 2004 Jun;134(2):93–119. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15153791.

[15] Kovats S, Main EK, Librach C, Stubblebine M, Fisher SJ, DeMars R. A class I antigen, HLA-G, expressed in human trophoblasts. Science (New York, NY). 1990 Apr;248(4952):220–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2326636.

[16] Simmons DG, Cross JC. Determinants of trophoblast lineage and cell subtype specification in the mouse placenta. Developmental biology. 2005 Aug;284(1):12–24. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15963972.

[17] Red-Horse K, Zhou Y. Trophoblast differentiation during embryo implantation and formation of the maternal-fetal interface. Journal of Clinical Investigation. 2004;114(6). Available from: http://www.jci.org/cgi/content/abstract/114/6/744.

[18] Aplin JD, Kimber SJ. Trophoblast-uterine interactions at implantation. Reproductive biology and endocrinology. 2004 Jul;2:48. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=471567&tool=pmcentrez&rendertype=abstract.

[19] Aplin JD. Embryo implantation: the molecular mechanism remains elusive. Reproductive biomedicine online. 2007 Jan;14 Spec No:49–55. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/20483399.

[20] Cross JC, Werb Z, Fisher SJ. Implantation and the placenta: key pieces of the development puzzle. Science (New York, NY). 1994 Dec;266(5190):1508–18. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7985020.

[21] Tanaka S. Promotion of Trophoblast Stem Cell Proliferation by FGF4. Science. 1998 Dec;282(5396):2072–2075. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.282.5396.2072.

[22] Roberts RM, Fisher SJ. Trophoblast stem cells. Biology of reproduction. 2011 Mar;84(3):412–21. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3043125&tool=pmcentrez&rendertype=abstract.

[23] Gardner RL. Origin and differentiation of extraembryonic tissues in the mouse. International review of experimental pathology. 1983 Jan;24:63–133. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6302028.

[24] Rossant J, Vijh KM. Ability of outside cells from preimplantation mouse embryos to form inner cell mass derivatives. Developmental biology. 1980 May;76(2):475–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6893035.

[25] Handyside AH. Time of commitment of inside cells isolated from preimplantation mouse embryos. Journal of embryology and experimental morphology. 1978 Jun;45:37–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/353216.

[26] Kunath T, Strumpf D, Rossant J. Early trophoblast determination and stem cell maintenance in the mouse–a review. Placenta. 2004 Apr;25 Suppl A:S32–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15033304.

[27] Gardner RL, Papaioannou VE, Barton SC. Origin of the ectoplacental cone and secondary giant cells in mouse blastocysts reconstituted from isolated trophoblast and inner cell mass. Journal of embryology and experimental

morphology. 1973 Dec;30(3):561–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/4772385.

[28] Roberts RM, Ezashi T, Das P. Trophoblast gene expression: transcription factors in the specification of early trophoblast. Reproductive biology and endocrinology : RB&E. 2004 Jul;2:47. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=471566&tool=pmcentrez&rendertype=abstract.

[29] Hu D, Cross JC. Development and function of trophoblast giant cells in the rodent placenta. The International journal of developmental biology. 2010 Jan;54(2-3):341–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19876834.

[30] James JL, Carter aM, Chamley LW. Human placentation from nidation to 5 weeks of gestation. Part I: What do we know about formative placental development following implantation? Placenta. 2012 May;33(5):327–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22374510.

[31] Cross JC, Simmons DG, Watson ED. Chorioallantoic morphogenesis and formation of the placental villous tree. Annals of the New York Academy of Sciences. 2003 May;995:84–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12814941.

[32] Sapin V, Dollé P, Hindelang C, Kastner P, Chambon P. Defects of the chorioallantoic placenta in mouse RXRalpha null fetuses. Developmental biology. 1997 Nov;191(1):29–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9356169.

[33] Adamson SL, Lu Y, Whiteley KJ, Holmyard D, Hemberger M, Pfarrer C, et al. Interactions between Trophoblast Cells and the Maternal and Fetal Circulation in the Mouse Placenta. Developmental Biology. 2002 Oct;250(2):358–373. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0012160602907736.

[34] Orwig K, Wolfe MW, Cohick C, Dai G. Trophoblast-specific regulation of endocrine-related genes. Placenta. 1998;p. 65–85. Available from: http://www.sciencedirect.com/science/article/pii/S0143400498800076.

[35] Simmons DG, Fortier AL, Cross JC. Diverse subtypes and developmental origins of trophoblast giant cells in the mouse placenta. Developmental biology. 2007 Apr;304(2):567–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17289015.

[36] Zybina EV, Zybina TG. Polytene chromosomes in mammalian cells. International review of cytology. 1996 Jan;165:53–119. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8900957.

[37] Gardner RL, Davies TJ. Lack of coupling between onset of giant transformation and genome endoreduplication in the mural trophectoderm of the mouse blastocyst. The journal of experimental zoology. 1993 Jan;265(1):54–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8459230.

[38] Soloveva V, Linzer DIH. Differentiation of placental trophoblast giant cells requires downregulation of p53 and Rb. Placenta. 2004 Jan;25(1):29–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15013636.

[39] Bevilacqua EM, Abrahamsohn PA. Ultrastructure of trophoblast giant cell transformation during the invasive stage of implantation of the mouse embryo. Journal of morphology. 1988 Dec;198(3):341–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3221406.

[40] Carney EW, Prideaux V, Lye SJ, Rossant J. Progressive expression of trophoblast-specific genes during formation of mouse trophoblast giant cells in vitro. Molecular reproduction and development. 1993 Apr;34(4):357–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8471259.

[41] Zybina EV, Zybina TG, Stein GI. Trophoblast cell invasiveness and capability for the cell and genome reproduction in rat placenta. Early pregnancy (Online).

2000 Jan;4(1):39–57. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/11719821.

[42] Enders AC, Welsh AO. Structural interactions of trophoblast and uterus during hemochorial placenta formation. The journal of experimental zoology. 1993 Sep;266(6):578–87. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/8371099.

[43] Pijnenborg R, Robertson WB, Brosens I, Dixon G. Review article: trophoblast invasion and the establishment of haemochorial placentation in man and laboratory animals. Placenta. 1981;2(1):71–91. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/7010344.

[44] Hemberger M, Hughes M, Cross JC. Trophoblast stem cells differentiate in vitro into invasive trophoblast giant cells. Developmental biology. 2004 Jul;271(2):362–71. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/15223340.

[45] Zechner U, Hemberger M, Constância M, Orth A, Dragatsis I, Lüttges A, et al. Proliferation and growth factor expression in abnormally enlarged placentas of mouse interspecific hybrids. Developmental dynamics : an official publication of the American Association of Anatomists. 2002 Jun;224(2):125–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12112466.

[46] Talamantes F. Structure and regulation of secretion of mouse placental lactogens. Progress in clinical and biological research. 1990 Jan;342:81–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2200057.

[47] Cross JC, Anson-Cartwright L, Scott IC. Transcription factors underlying the development and endocrine functions of the placenta. Recent progress in hormone research. 2002 Jan;57:221–34. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/12017545.

[48] Iwatsuki K, Shinozaki M, Sun W, Yagi S, Tanaka S, Shiota K. A novel secretory protein produced by rat spongiotrophoblast. Biology of reproduction. 2000

May;62(5):1352–9. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/10775187.

[49] Wynne F, Ball M, McLellan AS. Mouse pregnancy-specific glycoproteins: tissue-specific expression and evidence of association with maternal vasculature. Reproduction. 2006;Available from:
http://www.reproduction-online.org/content/131/4/721.short.

[50] Guzman-Ayala M, Ben-Haim N, Beck S, Constam DB. Nodal protein processing and fibroblast growth factor 4 synergize to maintain a trophoblast stem cell microenvironment. Proceedings of the National Academy of Sciences of the United States of America. 2004 Nov;101(44):15656–60. Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=524845&tool=pmcentrez&rendertype=abstract.

[51] Hemberger M, Dean W, Reik W. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. Nature reviews Molecular cell biology. 2009 Aug;10(8):526–37. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/19603040.

[52] Ralston A, Rossant J. How signaling promotes stem cell survival: trophoblast stem cells and Shp2. Developmental cell. 2006 Mar;10(3):275–6. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/16516829.

[53] Uy GD, Downs KM, Gardner RL. Inhibition of trophoblast stem cell potential in chorionic ectoderm coincides with occlusion of the ectoplacental cavity in the mouse. Development (Cambridge, England). 2002 Aug;129(16):3913–24. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12135928.

[54] Hunt CV, Avery GB. The development and proliferation of the trophoblast from ectopic mouse embryo allografts of increasing gestational age. Journal of reproduction and fertility. 1976 Mar;46(2):305–11. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/1255559.

[55] Senner CE, Hemberger M. Regulation of early trophoblast differentiation -

lessons from the mouse. Placenta. 2010 Nov;31(11):944–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20797785.

[56] Feldman B, Poueymirou W, Papaioannou VE, DeChiara TM, Goldfarb M. Requirement of FGF-4 for postimplantation mouse development. Science (New York, NY). 1995 Jan;267(5195):246–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7809630.

[57] Niswander L, Martin GR. Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. Development (Cambridge, England). 1992 Mar;114(3):755–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1618140.

[58] Rappolee DA, Basilico C, Patel Y, Werb Z. Expression and function of FGF-4 in peri-implantation development in mouse embryos. Development (Cambridge, England). 1994 Aug;120(8):2259–69. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7925026.

[59] Arman E, Haffner-Krausz R, Chen Y, Heath JK, Lonai P. Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. Proceedings of the National Academy of Sciences of the United States of America. 1998 Apr;95(9):5082–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=20217&tool=pmcentrez&rendertype=abstract.

[60] Ma GT, Soloveva V, Tzeng SJ, Lowe La, Pfendler KC, Iannaccone PM, et al. Nodal regulates trophoblast differentiation and placental development. Developmental biology. 2001 Aug;236(1):124–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11456449.

[61] Erlebacher A, Price Ka, Glimcher LH. Maintenance of mouse trophoblast stem cell proliferation by TGF-beta/activin. Developmental biology. 2004 Nov;275(1):158–69. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15464579.

[62] Guillemot F, Nagy A, Auerbach A, Rossant J, Joyner AL. Essential role of Mash-2 in extraembryonic development. Nature. 1994 Sep;371(6495):333–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8090202.

[63] Scott IC, Anson-Cartwright L, Riley P, Reda D, Cross JC. The HAND1 basic helix-loop-helix transcription factor regulates trophoblast differentiation via multiple mechanisms. Molecular and cellular biology. 2000 Jan;20(2):530–41. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=85124&tool=pmcentrez&rendertype=abstract.

[64] Hattori N, Nishino K, Ko YG, Hattori N, Ohgane J, Tanaka S, et al. Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. The journal of biological chemistry. 2004 Apr;279(17):17063–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14761969.

[65] Liu L, Roberts RM. Silencing of the gene for the beta subunit of human chorionic gonadotropin by the embryonic transcription factor Oct-3/4. The Journal of biological chemistry. 1996 Jul;271(28):16683–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8663260.

[66] Liu L, Leaman D, Villalta M, Roberts RM. Silencing of the gene for the alpha-subunit of human chorionic gonadotropin by the embryonic transcription factor Oct-3/4. Molecular endocrinology (Baltimore, Md). 1997 Oct;11(11):1651–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9328347.

[67] Ezashi T, Ghosh D, Roberts RM. Repression of Ets-2-induced transactivation of the tau interferon promoter by Oct-4. Molecular and cellular biology. 2001 Dec;21(23):7883–91. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=99954&tool=pmcentrez&rendertype=abstract.

[68] Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & development. 2003 Jan;17(1):126–40. Available from:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=195970&tool=pmcentrez&rendertype=abstract.

[69] Okuda A, Fukushima A, Nishimoto M, Orimo A, Yamagishi T, Nabeshima Y, et al. UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. The EMBO journal. 1998 Apr;17(7):2019–32. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1170547&tool=pmcentrez&rendertype=abstract.

[70] Kopp JL, Ormsbee BD, Desler M, Rizzino A. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem cells (Dayton, Ohio). 2008 Apr;26(4):903–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18238855.

[71] Donnison M, Beaton A, Davey HW, Broadhurst R, L'Huillier P, Pfeffer PL. Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. Development (Cambridge, England). 2005 May;132(10):2299–308. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15829518.

[72] Ng RKR, Dean W, Dawson C, Lucifero D, Madeja Z, Reik W, et al. Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. Nature cell biology. 2008 Nov;10(11):1280–90. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2635539&tool=pmcentrez&rendertype=abstract http://www.nature.com/ncb/journal/vaop/ncurrent/full/ncb1786.html http://www.ncbi.nlm.nih.gov/pubmed/18836439.

[73] Murre C, Voronova A, Baltimore D. B-cell- and myocyte-specific E2-box-binding factors contain E12/E47-like subunits. Molecular and cellular biology. 1991 Feb;11(2):1156–60. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=359799&tool=pmcentrez&rendertype=abstract.

[74] Riley P, Anson-Cartwright L, Cross JC. The Hand1 bHLH transcription factor

is essential for placentation and cardiac morphogenesis. Nature genetics. 1998 Mar;18(3):271–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9500551.

[75] Hughes M, Dobric N, Scott IC, Su L, Starovic M, St-Pierre B, et al. The Hand1, Stra13 and Gcm1 transcription factors override FGF signaling to promote terminal differentiation of trophoblast stem cells. Developmental biology. 2004 Jul;271(1):26–37. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15196947.

[76] Boudjelal M, Taneja R, Matsubara S, Bouillet P, Dolle P, Chambon P. Overexpression of Stra13, a novel retinoic acid-inducible gene of the basic helix-loop-helix family, inhibits mesodermal and promotes neuronal differentiation of P19 cells. Genes & Development. 1997 Aug;11(16):2052–2065. Available from: http://www.genesdev.org/cgi/doi/10.1101/gad.11.16.2052.

[77] Kraut N, Snider L, Chen CM, Tapscott SJ, Groudine M. Requirement of the mouse I-mfa gene for placental development and skeletal patterning. The EMBO journal. 1998 Nov;17(21):6276–88. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1170953&tool=pmcentrez&rendertype=abstract.

[78] Auman HJ, Nottoli T, Lakiza O, Winger Q, Donaldson S, Williams T. Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. Development (Cambridge, England). 2002 Jun;129(11):2733–47. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12015300.

[79] Yagi R, Kohn MJ, Karavanova I, Kaneko KJ, Vullhorst D, DePamphilis ML, et al. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. Development (Cambridge, England). 2007 Nov;134(21):3827–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17913785.

[80] Strumpf D, Mao CA, Yamanaka Y, Ralston A, Chawengsaksophak K, Beck F, et al. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development (Cambridge, England). 2005 May;132(9):2093–102. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15788452.

[81] Ma GT, Roth ME, Groskopf JC, Tsai FY, Orkin SH, Grosveld F, et al. GATA-2 and GATA-3 regulate trophoblast-specific gene expression in vivo. Development (Cambridge, England). 1997 Feb;124(4):907–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9043071.

[82] Ray S, Dutta D, Rumi MaK, Kent LN, Soares MJ, Paul S. Context-dependent function of regulatory elements and a switch in chromatin occupancy between GATA3 and GATA2 regulate Gata2 transcription during trophoblast differentiation. The Journal of biological chemistry. 2009 Feb;284(8):4978–88. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2643515&tool=pmcentrez&rendertype=abstract.

[83] Takeda K, Noguchi K, Shi W, Tanaka T, Matsumoto M, Yoshida N, et al. Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. Proceedings of the National Academy of Sciences of the United States of America. 1997 Apr;94(8):3801–4. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=20521&tool=pmcentrez&rendertype=abstract.

[84] Knöfler M. Critical growth factors and signalling pathways controlling human trophoblast invasion. The International journal of developmental biology. 2010;54:269–280. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/pmc2974212/.

[85] White LJ, Declercq W, Arfuso F, Charles AK, Dharmarajan AM. Function of caspase-14 in trophoblast differentiation. Reproductive biology and endocrinology : RB&E. 2009 Jan;7:98. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=

2753366&tool=pmcentrez&rendertype=abstract.

[86] Blanchon L, Bocco JL, Gallot D, Gachon aM, Lémery D, Déchelotte P, et al. Co-localization of KLF6 and KLF4 with pregnancy-specific glycoproteins during human placenta development. Mechanisms of development. 2001 Jul;105(1-2):185–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11429296.

[87] Tompers DM, Foreman RK, Wang Q, Kumanova M, Labosky Pa. Foxd3 is required in the trophoblast progenitor cell lineage of the mouse embryo. Developmental biology. 2005 Sep;285(1):126–37. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16039639.

[88] Liu Y, Labosky Pa. Regulation of embryonic stem cell self-renewal and pluripotency by Foxd3. Stem cells (Dayton, Ohio). 2008 Oct;26(10):2475–84. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2658636&tool=pmcentrez&rendertype=abstract.

[89] Douglas GC, VandeVoort Ca, Kumar P, Chang TC, Golos TG. Trophoblast stem cells: models for investigating trophectoderm differentiation and placental development. Endocrine reviews. 2009 May;30(3):228–40. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2726840&tool=pmcentrez&rendertype=abstract.

[90] Westerman BA, Chhatta A, Poutsma A, van Vegchel T, Oudejans CBM. NEUROD1 acts in vitro as an upstream regulator of NEUROD2 in trophoblast cells. Biochimica et biophysica acta. 2004 Jan;1676(1):96–103. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14732494.

[91] Basyuk E, Cross JC, Corbin J, Nakayama H, Hunter P, Nait-Oumesmar B, et al. Murine Gcm1 gene is expressed in a subset of placental trophoblast cells. Developmental dynamics : an official publication of the American Association of Anatomists. 1999 Apr;214(4):303–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10213386.

[92] Schreiber J, Riethmacher-Sonnenberg E, Riethmacher D, Tuerk EE, Enderich J, Bösl MR, et al. Placental failure in mice lacking the mammalian homolog of glial cells missing, GCMa. Molecular and cellular biology. 2000 Apr;20(7):2466–74. Available from: http://mcb.asm.org/content/20/7/2466.short http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=85439&tool=pmcentrez&rendertype=abstract.

[93] Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, et al. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell. 2005 Dec;123(5):917–29. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16325584.

[94] Kuijk EW, Du Puy L, Van Tol HTA, Oei CHY, Haagsman HP, Colenbrander B, et al. Differences in early lineage segregation between mammals. Developmental dynamics : an official publication of the American Association of Anatomists. 2008 Apr;237(4):918–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18330925.

[95] Nishioka N, Yamamoto S, Kiyonari H, Sato H, Sawada A, Ota M, et al. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. Mechanisms of development. 2008;125(3-4):270–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18083014.

[96] Russ AP, Wattler S, Colledge WH, Aparicio Sa, Carlton MB, Pearce JJ, et al. Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature. 2000 Mar;404(6773):95–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10716450.

[97] Cheng YH, Aronow BJ, Hossain S, Trapnell B, Kong S, Handwerger S. Critical role for transcription factor AP-2alpha in human trophoblast differentiation. Physiological genomics. 2004 Jun;18(1):99–107. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15039486.

[98] Richardson BD, Langland RA, Bachurski CJ, Richards RG, Kessler CA, Cheng

YH, et al. Activator protein-2 regulates human placental lactogen gene expression. Molecular and cellular endocrinology. 2000 Feb;160(1-2):183–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10715552.

[99] Kuckenberg P, Buhl S, Woynecki T, van Fürden B, Tolkunova E, Seiffe F, et al. The transcription factor TCFAP2C/AP-2gamma cooperates with CDX2 to maintain trophectoderm formation. Molecular and cellular biology. 2010 Jul;30(13):3310–20. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2897582&tool=pmcentrez&rendertype=abstract.

[100] Ralston A, Cox BJ, Nishioka N, Sasaki H, Chea E, Rugg-Gunn P, et al. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development (Cambridge, England). 2010 Feb;137(3):395–403. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20081188.

[101] 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. Developmental biology. 2001 Jul;235(2):422–32. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11437448.

[102] Loregger T. Regulatory Transcription Factors Controlling Function and Differentiation of Human Trophoblast - A Review. Placenta. 2003 Apr;24:S104–S110. Available from: http://linkinghub.elsevier.com/retrieve/pii/S014340040290929X.

[103] Fitzgerald JS, Busch S, Wengenmayer T, Foerster K, de la Motte T, Poehlmann TG, et al. Signal transduction in trophoblast invasion. Chemical immunology and allergy. 2005 Jan;88:181–99. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16129946.

[104] Brosens IA, Robertson WB, Dixon HG. The role of the spiral arteries in the pathogenesis of preeclampsia. Obstetrics and gynecology annual. 1972 Jan;1:177–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/4669123.

[105] Brümmendorf T, Rathjen FG. Cell adhesion molecules. 1: immunoglobulin superfamily. Protein profile. 1994 Jan;1(9):951–1058. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8528906.

[106] Rudert F, Saunders AM, Rebstock S, Thompson JA, Zimmermann W. Characterization of murine carcinoembryonic antigen gene family members. Mammalian genome : official journal of the International Mammalian Genome Society. 1992 Jan;3(5):262–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1638085.

[107] Ohta T. Evolution of gene families. Gene. 2000 Dec;259(1-2):45–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11163960.

[108] Louis EJ. Evolutionary genetics: making the most of redundancy. Nature. 2007 Oct;449(7163):673–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17928851.

[109] Kammerer R, Zimmermann W. Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families. BMC biology. 2010 Jan;8:12. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2832619&tool=pmcentrez&rendertype=abstract.

[110] Haig D. Genetic Conflicts in Human Pregnancy. The Quarterly Review of Biology. 1993 Dec;68(4):495. Available from: http://www.journals.uchicago.edu/doi/abs/10.1086/418300.

[111] Chuong EB, Tong W, Hoekstra HE. Maternal-fetal conflict: rapidly evolving proteins in the rodent placenta. Molecular biology and evolution. 2010 Jun;27(6):1221–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20123797.

[112] Hammarström S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. Seminars

in cancer biology. 1999 Apr;9(2):67–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10202129.

[113] Kuespert K, Pils S, Hauck CR. CEACAMs: their role in physiology and pathophysiology. Current opinion in cell biology. 2006 Oct;18(5):565–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16919437.

[114] Beauchemin N, Benchimol S, Cournoyer D, Fuks A, Stanners CP. Isolation and characterization of full-length functional cDNA clones for human carcinoembryonic antigen. Molecular and cellular biology. 1987 Sep;7(9):3221–30. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=367958&tool=pmcentrez&rendertype=abstract.

[115] Oikawa S, Nakazato H, Kosaki G. Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequence. Biochemical and biophysical research communications. 1987 Jan;142(2):511–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3814146.

[116] Zheng QX, Tease LA, Shupert WL, Chan WY. Characterization of cDNAs of the human pregnancy-specific beta 1-glycoprotein family, a new subfamily of the immunoglobulin gene superfamily. Biochemistry. 1990 Mar;29(11):2845–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2346748.

[117] Nagaishi T, Chen Z, Chen L, Iijima H, Nakajima a, Blumberg RS. CEACAM1 and the regulation of mucosal inflammation. Mucosal immunology. 2008 Nov;1 Suppl 1(November):S39–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19079227.

[118] Coutelier JP, Godfraind C, Dveksler GS, Wysocka M, Cardellichio CB, Noël H, et al. B lymphocyte and macrophage expression of carcinoembryonic antigen-related adhesion molecules that serve as receptors for murine coronavirus. European journal of immunology. 1994 Jun;24(6):1383–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8206098.

[119] Prall F, Nollau P, Neumaier M, Haubeck HD, Drzeniek Z, Helmchen U, et al.

CD66a (BGP), an adhesion molecule of the carcinoembryonic antigen family, is expressed in epithelium, endothelium, and myeloid cells in a wide range of normal human tissues. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society. 1996 Jan;44(1):35–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8543780.

[120] Kammerer R, Stober D, Singer BB, Obrink B, Reimann J. Carcinoembryonic antigen-related cell adhesion molecule 1 on murine dendritic cells is a potent regulator of T cell stimulation. Journal of immunology (Baltimore, Md : 1950). 2001 Jun;166(11):6537–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11359805.

[121] Gray-Owen SD, Blumberg RS. CEACAM1: contact-dependent control of immunity. Nature reviews Immunology. 2006 Jun;6(6):433–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16724098.

[122] Benchimol S, Fuks A, Jothy S, Beauchemin N, Shirota K, Stanners CP. Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule. Cell. 1989 Apr;57(2):327–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2702691.

[123] Taheri M, Saragovi U, Fuks A, Makkerh J, Mort J, Stanners CP. Self recognition in the Ig superfamily. Identification of precise subdomains in carcinoembryonic antigen required for intercellular adhesion. The Journal of biological chemistry. 2000 Sep;275(35):26935–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10864933.

[124] Tan K, Zelus BD, Meijers R, Liu Jh, Bergelson JM, Duke N, et al. Crystal structure of murine sCEACAM1a[1,4]: a coronavirus receptor in the CEA family. The EMBO journal. 2002 May;21(9):2076–86. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=125375&tool=pmcentrez&rendertype=abstract.

[125] Obrink B. CEA adhesion molecules: multifunctional proteins with signal-regulatory properties. Current opinion in cell biology. 1997

Oct;9(5):616–26. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/9330864.

[126] Muchová L, Vítek L, Jirsa M. [CEACAM1–a less well-known member of the family of carcinoembryonic antigens]. Casopís lékarˇů cˇeských. 2003 Jan;142(5):259–63. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/12920788.

[127] Lin T, Halbert S, Spellacy W. Measurement of pregnancy-associated plasma proteins during human gestation. Journal of Clinical Investigation. 1974;54(April):576–582. Available from:
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC301590/.

[128] Lee JN, Grudzinskas JG, Chard T. Circulating levels of pregnancy proteins in early and late pregnancy in relation to placental tissue concentration. British journal of obstetrics and gynaecology. 1979 Nov;86(11):888–90. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/315792.

[129] Kromer B, Finkenzeller D, Wessells J, Dveksler G, Thompson J, Zimmermann W. Coordinate expression of splice variants of the murine pregnancy-specific glycoprotein (PSG) gene family during placental development. European journal of biochemistry / FEBS. 1996 Dec;242(2):280–7. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/8973644.

[130] Rebstock S, Lucas K, Weiss M, Thompson J, Zimmermann W. Spatiotemporal expression of pregnancy-specific glycoprotein gene rnCGM1 in rat placenta. Developmental dynamics : an official publication of the American Association of Anatomists. 1993 Nov;198(3):171–81. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/8136522.

[131] Lei KJ, Sartwell aD, Pan CJ, Chou JY. Cloning and expression of genes encoding human pregnancy-specific glycoproteins. The Journal of biological chemistry. 1992 Aug;267(23):16371–8. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/1644821.

[132] Rettenberger G, Klett C, Zechner U, Kunz J, Vogel W, Hameister H. Visualization of the conservation of synteny between humans and pigs by heterologous chromosomal painting. Genomics. 1995 Mar;26(2):372–8. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/7601464`.

[133] Teglund S, Olsen A, Khan WN, Frangsmyr L, Hammarström S. The pregnancy-specific glycoprotein (PSG) gene cluster on human chromosome 19: fine structure of the 11 PSG genes and identification of 6 new genes forming a third subgroup within the carcinoembryonic antigen (CEA) family. Genomics. 1994;23(3):669–684. Available from: `http://www.sciencedirect.com/science/article/pii/S0888754384715564`.

[134] Rudert F, Zimmermann W, Thompson JA. Intra- and interspecies analyses of the carcinoembryonic antigen (CEA) gene family reveal independent evolution in primates and rodents. Journal of molecular evolution. 1989 Aug;29(2):126–34. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/2509715`.

[135] Zhou GQ, Baranov V, Zimmermann W, Grunert F, Erhard B, Mincheva-Nilsson L, et al. Highly specific monoclonal antibody demonstrates that pregnancy-specific glycoprotein (PSG) is limited to syncytiotrophoblast in human early and term placenta. Placenta. 1997 Sep;18(7):491–501. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/9290143`.

[136] Bohn H. [Detection and characterization of pregnancy proteins in the human placenta and their quantitative immunochemical determination in sera from pregnant women]. Archiv für Gynäkologie. 1971 Oct;210(4):440–57. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/5001318`.

[137] Gardner MO, Goldenberg RL, Cliver SP, Boots LR, Hoffman HJ. Maternal serum concentrations of human placental lactogen, estradiol and pregnancy specific beta 1-glycoprotein and fetal growth retardation. Acta obstetricia et gynecologica Scandinavica Supplement. 1997 Jan;165:56–8. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/9219458`.

[138] Gordon YB, Grudzinskas JG, Lewis JD, Jeffrey D, Letchworth AT. Circulating levels of pregnancy-specific beta1-glycoprotein and human placental lactogen in the third trimester of pregnancy: their relationship to parity, birth weight, and placental weight. British journal of obstetrics and gynaecology. 1977 Sep;84(9):642–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/911715.

[139] Gordon YP, Grudzinskas JG, Jeffrey D, Chard T. Concentrations of pregnancy-specific beta 1-glycoprotein in maternal blood in normal pregnancy and in intrauterine growth retardation. Lancet. 1977 Mar;1(8007):331–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/64859.

[140] Würz H, Geiger W, Künzig HJ, Jabs-Lehmann A, Bohn H, Lüben G. Radioimmunoassay of SP1 (pregnancy-specific beta1-glycoprotein) in maternal blood and in amniotic fluid normal and pathologic pregnancies. Journal of perinatal medicine. 1981 Jan;9(2):67–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6787188.

[141] Tamsen L. Pregnancy-specific beta 1-glycoprotein (SP1) levels measured by nephelometry in serum from women with vaginal bleeding in the first half of pregnancy. Acta obstetricia et gynecologica Scandinavica. 1984 Jan;63(4):311–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6611014.

[142] MacDonald DJ, Scott JM, Gemmell RS, Mack DS. A prospective study of three biochemical fetoplacental tests: serum human placental lactogen, pregnancy-specific beta 1-glycoprotein, and urinary estrogens, and their relationship to placental insufficiency. American journal of obstetrics and gynecology. 1983 Oct;147(4):430–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6605089.

[143] Grudzinskas JG, Gordon YB, Menabawey M, Lee JN, Wadsworth J, Chard T. Identification of high-risk pregnancy by the routine measurement of pregnancy-specific beta 1-glycoprotein. American journal of obstetrics and gynecology. 1983 Sep;147(1):10–2. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/6604456.

[144] Horne CH, Towler CM, Pugh-Humphreys RG, Thomson AW, Bohn H. Pregnancy specific beta1-glycoprotein–a product of the syncytiotrophoblast. Experientia. 1976 Sep;32(9):1197. Available from: http://www.ncbi.nlm.nih.gov/pubmed/971765.

[145] Hau J, Gidley-Baird AA, Westergaard JG, Teisner B. The effect on pregnancy of intrauterine administration of antibodies against two pregnancy-associated murine proteins: murine pregnancy-specific beta 1-glycoprotein and murine pregnancy-associated alpha 2-glycoprotein. Biomedica biochimica acta. 1985 Jan;44(7-8):1255–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3878707.

[146] McLellan AS, Zimmermann W, Moore T. Conservation of pregnancy-specific glycoprotein (PSG) N domains following independent expansions of the gene families in rodents and primates. BMC evolutionary biology. 2005 Jan;5:39. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1185527&tool=pmcentrez&rendertype=abstract.

[147] Teglund S, Zhou GQ, Hammarström S. Characterization of cDNA encoding novel pregnancy-specific glycoprotein variants. Biochemical and biophysical research communications. 1995 Jun;211(2):656–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7794280.

[148] McLellan AS, Fischer B, Dveksler G, Hori T, Wynne F, Ball M, et al. Structure and evolution of the mouse pregnancy-specific glycoprotein (Psg) gene locus. BMC genomics. 2005 Jan;6:4. Available from: http://w25.biomedcentral.com/1471-2164/6/4/ http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=546212&tool=pmcentrez&rendertype=abstract.

[149] Ruoslahti E, Pierschbacher MD. New perspectives in cell adhesion: RGD and integrins. Science (New York, NY). 1987 Oct;238(4826):491–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2821619.

[150] Kim JP, Zhang K, Chen JD, Wynn KC, Kramer RH, Woodley DT. Mechanism of human keratinocyte migration on fibronectin: unique roles of RGD site and integrins. Journal of cellular physiology. 1992 Jun;151(3):443–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1295896.

[151] Kim JP, Chen JD, Wilke MS, Schall TJ, Woodley DT. Human keratinocyte migration on type IV collagen. Roles of heparin-binding site and alpha 2 beta 1 integrin. Laboratory investigation; a journal of technical methods and pathology. 1994 Sep;71(3):401–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7933990.

[152] Kodelja V, Lucas K, Barnert S, von Kleist S, Thompson Ja, Zimmermann W. Identification of a carcinoembryonic antigen gene family in the rat. Analysis of the N-terminal domains reveals immunoglobulin-like, hypervariable regions. The Journal of biological chemistry. 1989 Apr;264(12):6906–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7851896 http://www.ncbi.nlm.nih.gov/pubmed/2708349 http://www.jbc.org/content/264/12/6906.short.

[153] McLane MA, Joerger T, Mahmoud A. Disintegrins in health and disease. Frontiers in bioscience : a journal and virtual library. 2008 Jan;13:6617–37. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18508683.

[154] Ruoslahti E, Pierschbacher MD. Arg-Gly-Asp: a versatile cell recognition signal. Cell. 1986 Feb;44(4):517–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2418980.

[155] Rutherfurd KJ, Chou JY, Mansfield BC. A motif in PSG11s mediates binding to a receptor on the surface of the promonocyte cell line THP-1. Molecular endocrinology (Baltimore, Md). 1995 Oct;9(10):1297–305. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8544838.

[156] Barbato O, Sousa NM, Klisch K, Clerget E, Debenedetti A, Barile VL, et al. Isolation of new pregnancy-associated glycoproteins from water buffalo (Bubalus bubalis) placenta by Vicia villosa affinity chromatography. Research

in veterinary science. 2008 Dec;85(3):457–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18308351.

[157] Szafranska B, Panasiewicz G, Majewska M. Biodiversity of multiple Pregnancy-Associated Glycoprotein (PAG) family: gene cloning and chorionic protein purification in domestic and wild eutherians (Placentalia)–a review. Reproduction, nutrition, development. 2006;46(5):481–502. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17107639.

[158] Motrán CC, Díaz FL, Gruppi A, Slavin D, Chatton B, Bocco JL. Human pregnancy-specific glycoprotein 1a (PSG1a) induces alternative activation in human and mouse monocytes and suppresses the accessory cell-dependent T cell proliferation. Journal of leukocyte biology. 2002 Sep;72(3):512–21. Available from: http://www.jleukbio.org/content/72/3/512.short http://www.ncbi.nlm.nih.gov/pubmed/12223519.

[159] Wessells J, Wessner DH, Parsells R, White K, Finkenzeller D, Zimmermann W, et al. Pregnancy specific glycoprotein 18 induces IL-10 expression in murine macrophages. European journal of immunology. 2000 Jul;30(7):1830–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10940872.

[160] Snyder SK, Wessner DH, Wessells J, Waterhouse R, Wahl LM, Zimmermann W, et al. Pregnancy-specific glycoproteins function as immunomodulators by inducing secretion of IL-10, IL-6 and TGF-beta1 by human monocytes. American journal of reproductive immunology (New York, NY : 1989). 2001 Apr;45(4):205–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11327547.

[161] Waterhouse R, Ha CT, Dveksler G. Murine CD9 is the receptor for pregnancy-specific glycoprotein 17. The Journal of experimental medicine. 2002;195(2):277–282. Available from: http://jem.rupress.org/content/195/2/277.abstract.

[162] Ellerman DA, Ha CT, Primakoff P, Myles DG, Dveksler GS. Direct binding of the ligand PSG17 to CD9 requires a CD9 site essential for sperm-egg fusion.

Molecular biology of the cell. 2003;14(12):5098. Available from: http://www.molbiolcell.org/cgi/content/abstract/14/12/5098.

[163] Ha CT, Waterhouse R, Wessells J, Wu JA, Dveksler GS. Binding of pregnancy-specific glycoprotein 17 to CD9 on macrophages induces secretion of IL-10, IL-6, PGE2, and TGF-beta1. Journal of leukocyte biology. 2005 Jun;77(6):948–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15772125.

[164] Ha CT, Waterhouse R, Warren J, Zimmermann W, Dveksler GS. N-glycosylation is required for binding of murine pregnancy-specific glycoproteins 17 and 19 to the receptor CD9. American journal of reproductive immunology (New York, NY : 1989). 2008 Mar;59(3):251–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18275518.

[165] Blois SM, Tirado-González I, Wu J, Barrientos G, Johnson B, Warren J, et al. Early expression of pregnancy-specific glycoprotein 22 (PSG22) by trophoblast cells modulates angiogenesis in mice. Biology of reproduction. 2012 Jun;86(6):191. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22423048.

[166] Martínez FF, Knubel CP, Sánchez MC, Cervi L, Motrán CC. Pregnancy-specific glycoprotein 1a activates dendritic cells to provide signals for Th17-, Th2-, and Treg-cell polarization. European journal of immunology. 2012 Jun;42(6):1573–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22678910.

[167] Marzi M, Vigano A, Trabattoni D, Villa ML, Salvaggio A, Clerici E, et al. Characterization of type 1 and type 2 cytokine production profile in physiologic and pathologic human pregnancy. Clinical and experimental immunology. 1996 Oct;106(1):127–33. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2200555&tool=pmcentrez&rendertype=abstract.

[168] Hill JA, Polgar K, Anderson DJ. T-helper 1-type immunity to trophoblast in

women with recurrent spontaneous abortion. JAMA : the journal of the American Medical Association. 1995 Jun;273(24):1933–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7783303.

[169] Moormann AM, Sullivan AD, Rochford RA, Chensue SW, Bock PJ, Nyirenda T, et al. Malaria and pregnancy: placental cytokine expression and its relationship to intrauterine growth retardation. The Journal of infectious diseases. 1999 Dec;180(6):1987–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10558956.

[170] Wu J, Johnson BL, Chen Y, Ha CT, Dveksler GS. Murine pregnancy-specific glycoprotein 23 induces the proangiogenic factors transforming-growth factor beta 1 and vascular endothelial growth factor a in cell types involved in vascular remodeling in pregnancy. Biology of reproduction. 2008 Dec;79(6):1054–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18753609 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2613688&tool=pmcentrez&rendertype=abstract.

[171] Lisboa FA, Warren J, Sulkowski GN, Aparicio M, David G, Zudaire E, et al. Pregnancy-specific glycoprotein 1 induces endothelial tubulogenesis through interaction with cell surface proteoglycans. Journal of Biological Chemistry. 2011;286(9):7577. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.5699&rep=rep1&type=pdf http://www.jbc.org/content/286/9/7577.short.

[172] Minagawa S, Nakabayashi K, Fujii M, Scherer SW, Ayusawa D. Early BrdU-responsive genes constitute a novel class of senescence-associated genes in human cells. Experimental cell research. 2005 Apr;304(2):552–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15748899.

[173] Endoh M, Kobayashi Y, Yamakami Y, Yonekura R, Fujii M, Ayusawa D. Coordinate expression of the human pregnancy-specific glycoprotein gene family during induced and replicative senescence. Biogerontology. 2009

Apr;10(2):213–21. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/18792801.

[174] Shanley DK, Kiely PA, Golla K, Allen S, Martin K, O'Riordan RT, et al. Pregnancy-specific glycoproteins bind integrin αIIbβ3 and inhibit the platelet-fibrinogen interaction. PloS one. 2013 Jan;8(2):e57491. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3585349&tool=pmcentrez&rendertype=abstract.

[175] Boucheix C, Rubinstein E. Tetraspanins. Cellular and molecular life sciences : CMLS. 2001 Aug;58(9):1189–205. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11577978.

[176] Sulkowski GN, Warren J, Ha CT, Dveksler GS. Characterization of receptors for murine pregnancy specific glycoproteins 17 and 23. Placenta. 2011 Jun;32(8):1–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3142296&tool=pmcentrez&rendertype=abstract http://www.ncbi.nlm.nih.gov/pubmed/21669460.

[177] Schaefer L, Schaefer RM. Proteoglycans: from structural compounds to signaling molecules. Cell and tissue research. 2010 Jan;339(1):237–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19513755.

[178] Kuberan B, Lech M, Borjigin J, Rosenberg RD. Light-induced 3-O-sulfotransferase expression alters pineal heparan sulfate fine structure. A surprising link to circadian rhythm. The Journal of biological chemistry. 2004 Feb;279(7):5053–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14630922.

[179] Beauvais DM, Ell BJ, McWhorter AR, Rapraeger AC. Syndecan-1 regulates alphavbeta3 and alphavbeta5 integrin activation during angiogenesis and is blocked by synstatin, a novel peptide inhibitor. The Journal of experimental medicine. 2009 Mar;206(3):691–705. Available from:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2699122&tool=pmcentrez&rendertype=abstract.

[180] Thompson J, Koumari R, Wagner K, Barnert S, Schleussner C, Schrewe H, et al. The human pregnancy-specific glycoprotein genes are tightly linked on the long arm of chromosome 19 and are coordinately expressed. Biochemical and biophysical research communications. 1990 Mar;167(2):848–59. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1690992.

[181] Camolotto S, Racca a, Rena V, Nores R, Patrito LC, Genti-Raimondi S, et al. Expression and transcriptional regulation of individual pregnancy-specific glycoprotein genes in differentiating trophoblast cells. Placenta. 2010 Apr;31(4):312–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20116096.

[182] Dimitriadou F, Phocas I, Mantzavinos T, Sarandakou A, Rizos D, Zourlas PA. Discordant secretion of pregnancy specific beta 1-glycoprotein and human chorionic gonadotropin by human pre-embryos cultured in vitro. Fertility and sterility. 1992 Mar;57(3):631–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1740210.

[183] Aronow BJ, Richardson BD, Handwerger S. Microarray analysis of trophoblast differentiation: gene expression reprogramming in key gene function categories. Physiological genomics. 2001 Jul;6(2):105–16. Available from: http://physiolgenomics.physiology.org/content/6/2/105.short http://www.ncbi.nlm.nih.gov/pubmed/11459926.

[184] Streydio C, Swillens S, Georges M, Szpirer C, Vassart G. Structure, evolution and chromosomal localization of the human pregnancy-specific beta 1-glycoprotein gene family. Genomics. 1990 Aug;7(4):661–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2387594.

[185] Wu SM, Bazar LS, Cohn ML, Cahill RA, Chan WY. Expression of pregnancy-specific beta 1-glycoprotein genes in hematopoietic cells. Molecular

and cellular biochemistry. 1993 May;122(2):147–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8232246.

[186] Searle F, Leake BA, Bagshawe KD, Dent J. Serum-SP1-pregnancy-specific-beta-glycoprotein in choriocarcinoma and other neoplastic disease. Lancet. 1978 Mar;1(8064):579–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/76123.

[187] Fagnart OC, Cambiaso CL, Lejeune MD, Noel G, Maisin H, Masson PL. Prognostic value of concentration of pregnancy-specific beta 1-glycoprotein (SP1) in serum of patients with breast cancer. International journal of cancer Journal international du cancer. 1985 Nov;36(5):541–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3876999.

[188] Horne CH, Reid IN, Milne GD. Prognostic significance of inappropriate production of pregnancy proteins by breast cancers. Lancet. 1976 Aug;2(7980):279–82. Available from: http://www.ncbi.nlm.nih.gov/pubmed/59853.

[189] Chamberlin ME, Lei KJ, Chou JY. Subtle differences in human pregnancy-specific glycoprotein gene promoters allow for differential expression. The Journal of biological chemistry. 1994 Jun;269(25):17152–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8006022.

[190] Salahshor S, Goncalves J, Chetty R, Gallinger S, Woodgett JR. Differential gene expression profile reveals deregulation of pregnancy specific beta1 glycoprotein 9 early during colorectal carcinogenesis. BMC cancer. 2005 Jan;5:66. Available from: http://www.biomedcentral.com/1471-2407/5/66/ http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1184062&tool=pmcentrez&rendertype=abstract.

[191] Ball M, McLellan AS, 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. 2004;325:103 – 113.

[192] Kawano K, Ebisawa M, Hase K, Fukuda S, Hijikata A, Kawano S, et al. Psg18 is specifically expressed in follicle-associated epithelium. Cell structure and function. 2007 Jan;32(2):115–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17984568.

[193] Phillips JM, Kuo IT, Richardson C, Weiss SR. A novel full-length isoform of murine pregnancy-specific glycoprotein 16 (psg16) is expressed in the brain but does not mediate murine coronavirus (MHV) entry. Journal of neurovirology. 2012 Apr;18(2):138–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22302612.

[194] Bocco JL, Panzetta-Dutari GM, Flury A, Patrito LC. Expression of pregnancy specific beta 1-glycoprotein gene in human placenta and hydatiform mole. Biochemistry international. 1989 May;18(5):999–1008. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2528957.

[195] Frangsmyr 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 biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2000;21(2):63–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10686536.

[196] Koritschoner NP, Bocco JL, Panzetta-Dutari GM, Dumur CI, Flury a, Patrito LC. A novel human zinc finger protein that interacts with the core promoter element of a TATA box-less gene. The Journal of biological chemistry. 1997 Apr;272(14):9573–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9083102.

[197] Blanchon L, Nores R, Gallot D, Marceau G, Borel V, Yang VW, et al. Activation of the human pregnancy-specific glycoprotein PSG-5 promoter by KLF4 and Sp1. Biochemical and biophysical research communications. 2006 May;343(3):745–753. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16563348 http:

//www.sciencedirect.com/science/article/pii/S0006291X06005420.

[198] Racca AC, Camolotto S, Ridano ME, Bocco JL, Genti-Raimondi S, Panzetta-Dutari GM. Krüppel-Like Factor 6 Expression Changes during Trophoblast Syncytialization and Transactivates ßhCG and PSG Placental Genes. PloS one. 2011 Jan;6(7):e22438. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3142166&tool=pmcentrez&rendertype=abstract.

[199] Nores R, Blanchon L, López-Díaz F, Bocco JL, Patrito LC, Sapin V, et al. 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. 2004 Jan;25(1):9–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15013634.

[200] López-Díaz F, Nores R, Panzetta-Dutari GM, Slavin D, Prieto C, Koritschoner NP, et al. RXRalpha regulates the pregnancy-specific glycoprotein 5 gene transcription through a functional retinoic acid responsive element. Placenta. 2007;28(8-9):898–906. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17475324.

[201] Oikawa D, Akai R, Iwawaki T. Positive contribution of the IRE1alpha-XBP1 pathway to placental expression of CEA family genes. FEBS letters. 2010 Mar;584(5):1066–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20146926.

[202] Panzetta-Dutari GM, Bocco JL, Reimund B, Flury A, Patrito LC. Nucleotide sequence of a pregnancy-specific beta 1 glycoprotein gene family member. Identification of a functional promoter region and several putative regulatory sequences. Molecular biology reports. 1992 Sep;16(4):255–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1454058.

[203] Lassar AB, Davis RL, Wright WE, Kadesch T, Murre C, Voronova A, et al. Functional activity of myogenic HLH proteins requires hetero-oligomerization

with E12/E47-like proteins in vivo. Cell. 1991 Jul;66(2):305–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1649701.

[204] Mangelsdorf DJ, Ong ES, Dyck JA, Evans RM. Nuclear receptor that identifies a novel retinoic acid response pathway. Nature. 1990 May;345(6272):224–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2159111.

[205] Panzetta-Dutari GM, Koritschoner N. Transcription of genes encoding pregnancy-specific glycoproteins is regulated by negative promoter-selective elements. Biochemical. 2000;519:511–519. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1221279/.

[206] Yoshida H, Matsui T, Yamamoto A, Okada T, Mori K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. Cell. 2001 Dec;107(7):881–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11779464.

[207] Mattick JS, Amaral PP, Dinger ME, Mercer TR, Mehler MF. RNA regulation of epigenetic processes. BioEssays : news and reviews in molecular, cellular and developmental biology. 2009 Jan;31(1):51–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19154003.

[208] Cheung P, Lau P. Epigenetic regulation by histone methylation and histone variants. Molecular endocrinology (Baltimore, Md). 2005 Mar;19(3):563–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15677708.

[209] Zlatanova J, Leuba SH, van Holde K. Chromatin fiber structure: morphology, molecular determinants, structural transitions. Biophysical journal. 1998 May;74(5):2554–66. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1299597&tool=pmcentrez&rendertype=abstract.

[210] Bernstein E, Allis CD. RNA meets chromatin. Genes & development. 2005 Jul;19(14):1635–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16024654.

[211] Imhof A. Epigenetic regulators and histone modification. Briefings in functional genomics & proteomics. 2006 Sep;5(3):222–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16951415.

[212] Kouzarides T. Chromatin modifications and their function. Cell. 2007 Feb;128(4):693–705. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17320507.

[213] Nakao M. Epigenetics: interaction of DNA methylation and chromatin. Gene. 2001 Oct;278(1-2):25–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11707319.

[214] Kamakaka RT, Biggins S. Histone variants: deviants? Genes & development. 2005 Feb;19(3):295–310. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15687254.

[215] Zhang Z, Pugh BF. High-resolution genome-wide mapping of the primary structure of chromatin. Cell. 2011 Jan;144(2):175–86. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3061432&tool=pmcentrez&rendertype=abstract.

[216] Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, et al. Long noncoding RNAs with enhancer-like function in human cells. Cell. 2010 Oct;143(1):46–58. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20887892.

[217] Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. EMBO reports. 2001 Nov;2(11):986–91. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1084129&tool=pmcentrez&rendertype=abstract.

[218] Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. Science (New York, NY). 2005 Sep;309(5740):1559–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16141072.

[219] Ma H, Hao Y, Dong X, Gong Q, Chen J, Zhang J, et al. Molecular Mechanisms and Function Prediction of Long Noncoding RNA. The Scientific World Journal. 2012;2012(1):1–11. Available from: http://www.hindawi.com/journals/tswj/2012/541786/.

[220] Amaral PP, Mattick JS. Noncoding RNA in development. Mammalian genome : official journal of the International Mammalian Genome Society. 2008 Aug;19(7-8):454–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18839252.

[221] Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annual review of biochemistry. 2012 Jan;81:145–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22663078.

[222] Prasanth KV, Spector DL. Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. Genes & development. 2007 Jan;21(1):11–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17210785.

[223] Moran Va, Perera RJ, Khalil AM. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. Nucleic acids research. 2012 Aug;40(14):6391–400. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3413108&tool=pmcentrez&rendertype=abstract.

[224] Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009 Mar;458(7235):223–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2754849&tool=pmcentrez&rendertype=abstract.

[225] Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010 Aug;142(3):409–19. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2956184&tool=pmcentrez&rendertype=abstract.

[226] Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nature genetics. 2010 Dec;42(12):1113–7. Available from: `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3040650&tool=pmcentrez&rendertype=abstract`.

[227] Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nature genetics. 2011 Jul;43(7):621–9. Available from: `http://www.ncbi.nlm.nih.gov/pubmed/21642992`.

[228] Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proceedings of the National Academy of Sciences of the United States of America. 2009 Jul;106(28):11667–72. Available from: `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2704857&tool=pmcentrez&rendertype=abstract`.

[229] Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes & development. 2011 Sep;25(18):1915–27. Available from: `http://genesdev.cshlp.org/content/25/18/1915.short http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3185964&tool=pmcentrez&rendertype=abstract`.

[230] Redon S, Reichenbach P, Lingner J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. Nucleic acids research. 2010 Sep;38(17):5797–806. Available from: `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2943627&tool=pmcentrez&rendertype=abstract`.

[231] Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, et al. The

nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Molecular cell. 2010 Sep;39(6):925–38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20797886.

[232] Collins K. Physiological assembly and activity of human telomerase complexes. Mechanisms of ageing and development;129(1-2):91–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2323683&tool=pmcentrez&rendertype=abstract.

[233] Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Molecular cell. 2008 Oct;32(2):232–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18951091.

[234] Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science (New York, NY). 2008 Dec;322(5908):1717–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18988810.

[235] Plath N, Ohana O, Dammermann B, Errington ML, Schmitz D, Gross C, et al. Arc/Arg3.1 is essential for the consolidation of synaptic plasticity and memories. Neuron. 2006 Nov;52(3):437–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17088210.

[236] Chowdhury S, Shepherd JD, Okuno H, Lyford G, Petralia RS, Plath N, et al. Arc/Arg3.1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. Neuron. 2006 Nov;52(3):445–59. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1784006&tool=pmcentrez&rendertype=abstract.

[237] Shepherd JD, Rumbaugh G, Wu J, Chowdhury S, Plath N, Kuhl D, et al. Arc/Arg3.1 mediates homeostatic synaptic scaling of AMPA receptors. Neuron. 2006 Nov;52(3):475–84. Available from:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1764219&tool=pmcentrez&rendertype=abstract.

[238] Kawashima T, Okuno H, Nonaka M, Adachi-Morishima A, Kyo N, Okamura M, et al. Synaptic activity-responsive element in the Arc/Arg3.1 promoter essential for synapse-to-nucleus signaling in activated neurons. Proceedings of the National Academy of Sciences of the United States of America. 2009 Jan;106(1):316–21. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2629236&tool=pmcentrez&rendertype=abstract.

[239] Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010 May;465(7295):182–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3020079 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3020079&tool=pmcentrez&rendertype=abstract.

[240] Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature. 2011 Apr;472(7341):120–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21423168.

[241] Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular biology and evolution. 1993 May;10(3):512–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8336541.

[242] Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution. 1985 Jul;39(4):783. Available from: http://www.jstor.org/stable/2408678?origin=crossref.

[243] Dye F. Obtaining Early Mammalian Embryos. Tested studies for laboratory teaching. 1993;7:97–112. Available from: http://www.ableweb.org/volumes/vol-7/8-dye.pdf.

[244] Conner D. Mouse embryo fibroblast (MEF) feeder cell preparation. Current Protocols in Molecular Biology. 2001 May;Chapter 23:Unit 23.2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18265203 http://onlinelibrary.wiley.com/doi/10.1002/0471142727.mb2302s51/full.

[245] Oda M, Shiota K, Tanaka S. Trophoblast stem cells. Methods in enzymology. 2006 Jan;419:387–400. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17141063.

[246] Himeno E, Tanaka S, Kunath T. Isolation and manipulation of mouse trophoblast stem cells. Current protocols in stem cell biology. 2008 Oct;Chapter 1(October):Unit 1E.4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18972374.

[247] Soares MJ, Konno T, Alam SMK. The prolactin family: effectors of pregnancy-dependent adaptations. Trends in endocrinology and metabolism: TEM. 2007 Apr;18(3):114–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17324580.

[248] Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser aG, Catanese JJ, et al. Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. Genome research. 2000 Jan;10(1):116–28. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=310499&tool=pmcentrez&rendertype=abstract.

[249] Rubinson D, Dillon C, Kwiatkowski A. A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. Nature genetics. 2003 Mar;33(3):401–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12590264 http://www.nature.com/ng/journal/v33/n3/abs/ng1117.html.

[250] Ventura A, Meissner A. Cre-lox-regulated conditional RNA interference from transgenes. Proceedings of the National Academy of Sciences of the United States of America. 2004 Jul;101(28):10380–5. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=

478580&tool=pmcentrez&rendertype=abstract http://www.pnas.org/content/101/28/10380.short.

[251] Vandenbroucke II, Vandesompele J, Paepe aD, Messiaen L. Quantification of splice variants using real-time PCR. Nucleic acids research. 2001 Jul;29(13):E68–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=55792&tool=pmcentrez&rendertype=abstract.

[252] Goossens K, Van Soom A, Van Zeveren A, Favoreel H, Peelman LJ. Quantification of fibronectin 1 (FN1) splice variants, including two novel ones, and analysis of integrins as candidate FN1 receptors in bovine preimplantation embryos. BMC developmental biology. 2009 Jan;9:1. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2648952&tool=pmcentrez&rendertype=abstract.

[253] Prete MD, Vernal R, Dolznig H. Isolation of polysome-bound mRNA from solid tissues amenable for RT-PCR and profiling experiments. Rna. 2007;13(3):414–421. Available from: http://rnajournal.cshlp.org/content/13/3/414.short.

[254] Stone AB. a simplified method for preparing sucrose gradients. biochem j. 1973;137:117–118.

[255] Sinha AU, Meller J. Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. BMC bioinformatics. 2007 Jan;8:82. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1821339&tool=pmcentrez&rendertype=abstract.

[256] Beauchemin N, Draber P, Dveksler G, Gold P, Gray-Owen S, Grunert F, et al. Redefined nomenclature for members of the carcinoembryonic antigen family. Experimental cell research. 1999 Nov;252(2):243–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11501563?dopt=Abstract http://www.ncbi.nlm.nih.gov/pubmed/11501563.

[257] Mangelsdorf DJ, Borgmeyer U, Heyman Ra, Zhou JY, Ong ES, Oro aE, et al. Characterization of three RXR genes that mediate the action of 9-cis retinoic acid. Genes & Development. 1992 Mar;6(3):329–344. Available from: http://www.genesdev.org/cgi/doi/10.1101/gad.6.3.329.

[258] Mowat AM. Anatomical basis of tolerance and immunity to intestinal antigens. Nature reviews Immunology. 2003 Apr;3(4):331–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12669023.

[259] Müllner E, Garcia-Sanz J. Polysome gradients. Manual of Immunology Methods. 1997;p. 457–462. Available from: http://www.researchgate.net/publication/231637449_Chapter_7.7_Polysome_Gradients/file/9fcfd506ee5369746b.pdf.

[260] Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, et al. Modeling gene expression using chromatin features in various cellular contexts. Genome biology. 2012 Sep;13(9):R53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22950368.

[261] Thompson J, Zimmermann W. The carcinoembryonic antigen gene family: structure, expression and evolution. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 1988 Jan;9(2-3):63–83. Available from: http://content.karger.com/ProdukteDB/produkte.asp?Doi=217547 http://www.ncbi.nlm.nih.gov/pubmed/3041547.

[262] Letterio JJ, Roberts AB. Regulation of immune responses by TGF-beta. Annual review of immunology. 1998 Jan;16:137–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9597127.

[263] Yang L, Pang Y, Moses HL. TGF-beta and immune cells: an important regulatory axis in the tumor microenvironment and progression. Trends in immunology. 2010 Jun;31(6):220–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2891151&tool=pmcentrez&rendertype=abstract.

[264] Venkatesha S, Toporsian M, Lam C, Hanai Ji, Mammoto T, Kim YM, et al. Soluble endoglin contributes to the pathogenesis of preeclampsia. Nature medicine. 2006 Jun;12(6):642–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16751767.

[265] Ha CT, Waterhouse R. N-glycosylation is Required for Binding of Murine Pregnancy-Specific Glycoproteins 17 and 19 to the Receptor CD9. American Journal of .... 2008;p. 251–258. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0897.2007.00573.x/full.

[266] Hannon GJ. RNA interference. Nature. 2002 Jul;418(6894):244–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15965464 http://www.ncbi.nlm.nih.gov/pubmed/12110901.

[267] Kunath T, Gish G, Lickert H, Jones N, Pawson T, Rossant J. Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. Nature biotechnology. 2003 May;21(5):559–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12679785.

[268] Pan G, Li J, Zhou Y, Zheng H, Pei D. A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2006 Aug;20(10):1730–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16790525.

[269] Smale ST. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochimica et biophysica acta. 1997 Mar;1351(1-2):73–88. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9116046.