TITLE: Convergent evolution of pregnancy-specific glycoproteins in human and horse

AUTHORS: Aleksic, D; Blaschke, L; Missbach, S; Hänske, J; Weiss, W; Handler, J; Zimmermann, W; Cabrera-Sharp, V; Read, J E; De Mestre, A M; O'Riordan, R; Moore, T; Kammerer, R

JOURNAL TITLE: Reproduction

PUBLICATION DATE: September 2016

PUBLISHER: Bioscientifica

DOI: 10.1530/REP-16-0236

Trophoblast expression and function of pregnancy-specific glycoproteins (PSGs) in the horse

Denis Aleksic[1#], Lisa Blaschke[1#], Sophie Mißbach[1], Jana Hänske[1], Wiebke Weiß[1], Johannes Handler[2], Wolfgang Zimmermann[3],Victoria Cabrera-Sharp[4], Jordan E. Read[4], Amanda M de Mestre[4], Ronan O'Riordan[5], Tom Moore[5], Robert Kammerer[1]

[1]Institute of Immunology, Friedrich-Loeffler-Institut, Suedufer 10, Greifswald-Insel Riems, Germany. [2]Equine Center Bad Saarow, Clinic for Horses, Veterinary Faculty, Freie Universität Berlin, Bad Saarow, Germany. [3]Tumor Immunology Laboratory, LIFE Center, University Clinic, Ludwig-Maximilians-University, Munich, Germany. [4]Department of Comparative Biomedical Sciences, The Royal Veterinary College, University of London, London, United Kingdom. [5]School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland.

[#]D.A. and L.B. contributed equally to this work.

Address correspondence to:
Robert Kammerer
Friedrich-Loeffler-Institut
Bundesforschungsinstitut für Tiergesundheit
Federal Research Institute for Animal Health
Südufer 10
D -17493 Greifswald – Insel Riems, Germany
Email: Robert.Kammerer@fli.bund.de
Tel.: +49 383517 1600

Short title: Convergent evolution of PSGs

Abstract

Pregnancy-Specific Glycoproteins (PSGs) are secreted members of the carcinoembryonic antigen cell adhesion molecule (CEACAM) family, which are expressed by trophoblast cells. PSGs may modulate maternal immune and platelet responses during pregnancy. Until now PSGs are only found in species which have a highly invasive (hemochorial) placentation including humans, mice and rats. Surprisingly, analyzing the CEACAM gene family of the horse, which has a less invasive epitheliochorial placenta, we identified equine CEACAM family members that seem to be related to PSGs of rodents and primates. We identified seven genes that encode secreted PSG-like CEACAMs. Phylogenetic analyses indicate that they evolved independently from an equine CEACAM1-like ancestor rather than from a common PSG-like ancestor with rodents and primates. Significantly, expression of PSG-like genes (CEACAM44, CEACAM48, CEACAM49, CEACAM55) was found in trophoblast cells such as purified chorionic girdle cells and endometrial cup cells. Chorionic girdle cells are highly invasive trophoblast cells which invade the endometrium of the mare where they form endometrial cups andare in close contact with maternal immune cells. Therefore despite some fundamental differences the microenvironment of invasive equine trophoblast cells has striking similarities to the microenvironment of trophoblast cells in hemochorial placentas, suggesting that equine PSG-like CEACAMs and rodent and primate PSGs have undergone convergent evolution as an adaption to a highly specific microenvironment. This is supported by our finding that, similar to certain rodent and human PSGs, equine PSG-like CEACAM49 has anti-platelet activity. Our results have implications for understanding the evolution of PSGs and their functions in maternal-fetal interactions.

Introduction

The placenta has pivotal functions in facilitating nutrient exchange between mother and fetus and in protecting the fetus from the maternal immune system (Hemberger 2013). This implies that there may be common placenta-specific genes expressed in all placental mammals. However, the composition of placenta-specific genes appears to be highly species-specific, partly due to the expansion of different gene families in different mammalian lineages (Rawn & Cross 2008). Indeed, placentas are quite diverse among mammals in their structure, endocrine function and immunology. For example, the extent of invasion of the uterus by placental trophoblast cells can range from no invasion (as in epitheliochorial placentation) to very extensive invasion (as in hemochorial placentation) where fetal trophoblast cells are in direct contact with maternal blood (Moffett & Loke 2006).

In the highly invasive placentas of the human and laboratory animals such as mice, rats and rabbits, potent tolerance mechanisms are required to avoid deleterious maternal immune and thrombotic responses to the semiallogeneic fetus (von Rango 2008). Many mechanisms have been proposed to explain the maternal-fetal immune tolerance since Medawar and colleagues first discussed the idea of maternal-fetal tolerance (Trowsdale & Betz 2006, Schumacher *et al.* 2014).

One protein family which is thought to be involved in the regulation of the immune and platelet responses at the maternal-fetal interface is the Pregnancy-specific glycoprotein family (PSG). PSGs were originally isolated from the circulation of pregnant women (Bohn 1971). In humans, the concentration of PSGs in the bloodstream increases exponentially until term, thereby comprising the most abundant fetal proteins in maternal blood (Lin *et al.* 1974). PSGs are placentally secreted proteins and part of the carcinoembryonic antigen (CEA) family, which by itself is a member of the immunoglobulin superfamily. The CEA gene family comprises a group of "conserved" CEACAMs, and a larger group of CEACAM1 paralogs. The "conserved" CEACAMs include *CEACAM1* itself, *CEACAM16*, *CEACAM18*, *CEACAM19* and *CEACAM20* (Kammerer & Zimmermann 2010). The *PSG* genes belong to the CEA-related cell adhesion molecule 1 (*CEACAM1*) paralogs of the CEA gene

family. PSGs are secreted CEACAMs expressed in a predominantly trophoblast-specific manner.

PSGs are one of the most rapidly evolving protein families in humans (Zebhauser *et al.* 2005, Chuong *et al.* 2010). Probably as a consequence, PSGs differ considerably in structure between primates and rodents. Human PSGs are composed of one N-terminal immunoglobulin variable (IgV)-like domain (N domain) followed by two to three Ig constant (IgC)-like domains of two different types (named A and B), whereas rodent PSGs contain 2 - 9 consecutive N domains followed by one IgC-like domain (McLellan *et al.* 2005). In addition, the expansion of *PSG* genes took place in two different loci in higher primates and rodents (Kammerer & Zimmermann 2010). There exist 11 and 17 different *PSG* genes in humans and mice, respectively. Recently it was demonstrated that human PSGs are expressed in trophoblast cells due to the activation of various transcription factors including Sp1 (specificity protein 1), KLF4 (Krüppel-like transcription factor 4) and RXRα (retinoid X receptor alpha) (Camolotto *et al.* 2010, Hubert *et al.* 2010). If this is also the case for rodent PSGs is not known.

Low PSG levels in maternal circulation are associated with certain pregnancy complications such as intrauterine growth retardation, preeclampsia and spontaneous abortion (Towler *et al.* 1977, Karg *et al.* 1981, Silver *et al.* 1993, Arnold *et al.* 1999). This may be due to loss of the immunoregulatory and anti-thrombotic functions of PSGs. PSGs induce monocytes to synthesize anti-inflammatory cytokines and promote alternative macrophage activation which correlates with the shift from Th1- to Th2-mediated immunological responses (Wessells *et al.* 2000, Snyder *et al.* 2001, Motran *et al.* 2003). In addition elevated PSG levels are correlated with improved symptoms of rheumatoid arthritis (Fialova *et al.* 1991) and multiple sclerosis (Bebo & Dveksler 2005). PSGs also exhibit anti-thrombotic activity by binding integrin αIIbβ3 and inhibiting the platelet – fibrinogen interaction (Shanley *et al.* 2013). To our current knowledge, PSGs evolved exclusively in mammals with hemochorial placentation. In cattle (epitheliochorial placenta) and dogs (endotheliochorial placenta), PSGs did not evolve (Kammerer *et al.* 2004, Kammerer *et al.* 2007).

However, here we show that PSGs evolved in the horse. The horse has an epitheliochorial placenta,

which represents the least invasive form of placentation. However, a subpopulation of trophoblast cells (chorionic girdle cells) differentiates to a highly invasive phenotype (Allen *et al.* 1973, Lunn *et al.* 1997). Around day 35 post-ovulation these binucleated, well-differentiated, eCG-secreting cells invade the endometrium. and build the endometrial cups which reach their maximum size and hormone-secreting capacity between day 60 and 70 of gestation (Allen 2001). During invasion, chorionic girdle cells express conventional MHC class I antigens and are recognized but not initially destroyed by the maternal immune system (Donaldson *et al.* 1990, Noronha & Antczak 2010). The endometrial cups have a restricted lifespan and the cells die around days 80-100 of pregnancy, although the exact mechanisms that lead to their destruction remain poorly understood. In this report, we show that equine PSGs are expressed in Chorionic girdle cells as well as in endometrial cups and that equine PSGs have a similar functional activity to human and rodent PSGs. Future work is in progress to examine the immunomodulatory functions of equine PSGs.

Material and Methods

Cells and tissues

Different equine tissue samples including salivary gland, trachea, lung, kidney, renal pelvis, skin, fat, pancreas, vein, liver, spleen, lymph node, tongue, esophagus, stomach, duodenum, jejunum, ileum, cecum, colon, rectum, ovary, uterus, cervix, udder, oviduct, bladder, vulva were collected from freshly slaughtered healthy horses and flash-frozen in liquid nitrogen. Peripheral blood mononuclear cells (PBMCs) were isolated from blood, which was collected for diagnostic purpose, of healthy horses by density-gradient centrifugation through Ficoll-Paque 1.077 g/l (GE Healthcare, Freiburg, Germany).

Stimulation of PBMC with 200 U/ml human IL-2 for the indicated time was performed at a concentration of 5 x 10^5 cells/ml in RPMI-1640 supplemented with 10% fetal calf serum (FCS "Gold"; PAA Laboratories, Coelbe, Germany), 2 mM L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin, non-essential amino acids and 1 mM sodium pyruvate (GIBCO/Invitrogen, Karlsruhe, Germany). Endometrial cup tissue samples were obtained from a 65 days pregnant, sedated mare by transcervical hysteroscopy via an endoscopy biopsy forceps under visual control. The procedure was performed to

initiate the abortion of the unwanted fetus. Tissue samples were immediately transferred to RNAlater (Qiagen, Hilden, Germany). Relevant tissue collection was approved by animal use committee of local authorities (LALLF Rostock; 7221.3-2.1-011/13). Generation of tissue from earlier stages of pregnancy and isolation and culture of chorionic girdle trophoblast cells has been previously described (Cabrera-Sharp *et al.* 2014). In brief, mares aged 3–7 years were maintained at the Royal Veterinary College and animal care was performed in accordance with the Animals (Scientific Procedures) Act 1986 guidelines set by the Home Office and the Ethics Committee of the Royal Veterinary College. The reproductive cycle of 5 mares was manipulated, and pregnancies were established using semen from 3 stallions. Conceptuses were recovered by nonsurgical uterine lavage with established methods between days 31 and 34 of pregnancy. Conceptuses were microdissected into chorionic girdle, allantochorion, chorion, yolk sac, bilaminar omphalopleure, and fetus and tissue snap frozen in liquid nitrogen. To gain a pure population of trophoblast cells, strips of chorionic girdle were placed into DMEM, and the chorionic girdle trophoblast cells were gently removed from basement membrane and underlying avascular mesodermal cell layer and cultured as per published methods (de Mestre *et al.* 2008). Tissues from 4 independent pregnancies and trophoblast cells isolated from an additional 5 independent pregnancies were used with the gestational age and passage number (0-3) of the cultured cells indicated in the figure legend.

Identification and prediction of equine CEACAMs

In order to identify all genes of the equine CEA gene family cluster we screened both the whole genome shot gun sequence (WGS) database at NCBI (http://www.ncbi.nlm.nih.gov) and the ensemble genome build at (http://www.ensemble.org) by running respective BLAST search tools using nucleotide sequences of all CEACAM exons of human, mouse, dog and cattle. Exon/Intron structures were predicted manually.

Reverse transcription-polymerase chain reaction analysis

Total RNA extraction from tissue was performed using the RNeasy kit (Qiagen, Hilden, Germany). RNA was isolated from cultured trophoblast cells using RNAbee (Amsbio, Abingdon, United

Kingdom). RNAbee was added directly to the culture well and cells scraped before removal to an Eppendorf tube and RNA extracted as described by the manufacturer. 500ng to one microgram of total RNA was used for cDNA synthesis by reverse transcription (RT) using the AMV Reverse Transcriptase (Promega, Mannheim, Germany). The RT product was amplified by polymerase chain reaction (PCR) with Taq polymerase (Fermentas, thermos Scientific, VWR International Germany GmbH, Darmstadt, Germany). After an initial denaturation step at 95°C for 45 s, 35 PCR cycles (denaturation: 95°C, 30 s; annealing: 60°C, 1 min; extension: 72°C, 1.5 min) and a final extension step at 72°C for 15 min were performed. Primers were designed according to our sequence predictions based on sequences identified in various NCBI data bases and based on newly identified sequences in this work. All primers used are summarized in Table 1. Eight microliters of each PCR were analyzed by electrophoresis on a 1.8% agarose gel and visualized by ethidium bromide staining. Real Time quantitative Polymerase Chain Reaction (qPCR) was performed with Maxima SYBR Green/Fluorescein qPCR Master Mix (Thermo Scientific). An initial denaturation at 95°C for 10 min to activate Maxima Hot Start Taq DNA polymerase, was followed by 40 cycles of a two-step cycling amplification (95°C for 15 s, 60°C for 1 min). The relative abundance of transcripts was calculated in arbitrary units (AU) using the formula 2^{35-cq} (AU), where cq represents the quantification cycle.

cDNA cloning

Primers used for amplification of full-length cDNAs (CEACAM46[a] CEACAM46b CEACAM47, CEACAM49 CEACAM55) are summarized in table 1. These primers introduced *Hin*dIII or *Xho*I and *Xba*I or *Eco*RI restriction sites into the PCR products corresponding to the 5'- and 3'-ends of the mRNA, respectively. The full-length cDNA was digested with *Hind*III or *Xho*I and *Xba*I or *Eco*RI and cloned into the pRc/CMV or pEGFP-N3 expression vector. Plasmid DNA isolated from various clones were analyzed by PCR and sequencing. Nucleotide sequencing was performed with the BigDye Terminator Cycle Sequencing Kit (PE Applied Biosystems, Weiterstadt, Germany).

Phylogenetic analyses

Phylogenetic analyses based on nucleotide and amino acid sequences were conducted using MEGA

version 5. The Neighbor-Joining (NJ) method with bootstrap testing (100 or 1000 replicates) and Poisson correction was applied.

CEACAM49 and PSG1 protein production and platelet assay

Protein production and purification was done as described (Shanley *et al.* 2013) using transient transfection of pTT3CC49 and pTT3PSG1 vectors, respectively, into HEK-293 cells (Freestyle™ 293 cells). The pTT3CC49 (pTT3CEACAM49) construct was generated by PCR amplifying the CEACAM49 CDS from pRc/CMV-ecaCEACAM49 incorporating 5' *Eco*RI and 3' *Hin*dIII restriction sites using the primers, ecaCEACAM49For: 5'-gGAATTCaccatgcaatcaccctca and ecaCEACAM49Rev: 5'-cAAGCTtggctctgtaactgggg. The PCR amplified band was blunt cloned into PSTBlue1 and sub-cloned into pTT3-B using *Eco*RI and *Hin*dIII restriction enzymes.

To produce recombinant PSG1 and CEACAM49 proteins, endotoxin-free plasmid DNA was purified from bacterial cultures using the Endofree Plasmid Maxi Kit (Qiagen). 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 (Invitrogen) using Freestyle MAX reagent (Invitrogen). The Freestyle 293-F cells were grown in suspension in Freestyle 293 Expression Medium, by shaker culture, to a density of 1×10^6 cells per ml. The plasmid DNA was diluted in OptiPRO serum-free medium at a ratio of 1µg DNA in 20µL OptiPRO for every 1mL of cells. Freestyle MAX reagent was also diluted in OptiPRO at the same ratio (1µL Freestyle MAX reagent in 20µL OptiPRO per milliliter of cells). The diluted DNA and Freestyle MAX reagent were then combined, mixed gently, and incubated at room temperature (RT) for 20min. The mixture was added to the cell suspension and the cells were cultured for a further 72h. The culture was then centrifuged at 200g for 5min at RT to separate the protein-containing medium from the cells, and the medium was frozen in aliquots at −80°C.

Recombinant proteins were purified from cell culture medium using Qiagen Ni-NTA resin. Imidazole (Sigma) was added to the culture medium to a final concentration of 10mM to reduce non-specific binding. Ni-NTA resin was added to the medium at a ratio of 1ml resin suspension (corresponding to 0.5mL resin bed volume) to 100mL 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) and the resin was washed with wash buffer (500mM NaCl, 20mM NaH2PO4, pH6) until the absorbance at 260nm reduced to 0. Protein was then eluted from the column with increasing concentrations of imidazole in wash buffer, usually 4×1.5ml 50mM fractions, 5×1.5mL 200mM fractions, 4×1.5ml 300mM fractions and 3×1.5ml 500mM fractions. Fractions containing recombinant PSG1 or CEACAM49 were then pooled and passed through a new column containing 0.5ml Ni-NTA resin, the flow-through collected, and bound protein eluted with 1mL 50mM imidazole followed by three 1mL 200mM imidazole fractions. The flow-through and three 200mM fractions were pooled and concentrated to a volume of 4–6mL using a Millipore Amicon Ultra Ultracel 10K centrifugal filter (Millipore). The concentrate was then dialyzed against three changes of 2L of phosphate-buffered saline (PBS) at 4°C. The protein was then further concentrated to a volume of 1–2mL depending on the starting volume of culture medium. Protein was quantified by Bradford Assay or UV Spectroscopy, checked by polyacrylamide gel electrophoresis, tested for LPS contamination (Limulus Amebocyte Lysate QCL-1000; Cambrex BioScience, Germany), aliquoted, and frozen at −80°C. For size exclusion chromatography (SEC), 1–4mg purified recombinant protein was applied to a HiLoad 16/60 Superdex S200 prep grade column in a 1mL volume with a flow rate of 1mL/min using the Akta explorer system in PBS. Determination of 1mL fractions to pool was based on chromatogram peaks. Resultant protein solution was concentrated. TGF β 1 contamination level was determined by enzyme-linked immunosorbent assay (ELISA).

Preparation of human platelets

Platelets were collected into 0.15vol/vol acid-citrate dextrose (ACD; 75mM trisodium citrate, 124mM dextrose and 38mM citric acid) anticoagulant and washed. Briefly, blood was centrifuged at 150***g*** for 10min at RT. In order to avoid any contamination from the buffy coat, ~0.5mL of the platelet-rich suspension above the buffy coat layer was left behind in the centrifugation tube. Platelet-rich plasma (PRP) was then acidified to pH6.5 with ACD and prostaglandin E_1 (PGE_1, 1mM) was added. The platelets were pelleted by centrifugation at 750***g*** for 10min at RT. The supernatant was removed and the platelet pellet was gently resuspended in 130mM NaCl, 3mM KCl, 10mM trisodium citrate, 9mM $NaHCO_3$, 6mM dextrose, 0.9mM $MgCl_2$, 0.81mM KH_2PO_4 and 10mM Tris pH7.4 (JNL buffer). Platelet

count was adjusted to 3×10^8per mL using a Sysmex XE K-1000 counter (Toa Medical Electronics Co. Ltd, Kobe, Japan). Platelets were allowed to stand at RT for 45min to let PGE_1dissipate. Calcium chloride ($CaCl_2$, 1.8mM) was added to platelets immediately before use.

Platelet integrin binding assay

Platelet integrin interaction assay was done as described (Shanley *et al.* 2013). Briefly, 10µL of 2.5mg/mL Oregon green-labeled Fibrinogen (OgFg; Invitrogen) was added to 20µL aliquots of washed platelet suspension along with the indicated concentrations of PSG1 or CEACAM49 in successive tubes. Dose–response curves were produced for all proteins using serial dilutions of 1:2 in the range 200 to 12.5µg/mL. All experiments were run in duplicate at least three times. The platelet suspension was vortexed and allowed to stand at RT for 10min before the addition of 3.4µM thrombin receptor activating peptide (TRAP, Bachem, UK), a dose known to produce a 50% maximal response as measured by aggregometry in pilot studies. Assay tubes were incubated at RT for a further 10min. The reaction was stopped by addition of 1mL ice-cold buffer. In a separate series of experiments, 250nM U46619 (Tocris Bioscience, UK), 10µM ADP (Bio/Data Corporation, UK), or 25µM epinephrine (Bio/Data Corporation, UK) were used to activate platelets. The association of OgFg with platelets was detected using a fluorescence-activated cell sorter (Becton Dickinson, Franklin Lakes, NJ, USA). Data acquisition and analysis were performed with the Cell Quest program. Platelet populations were gated, and histograms of mean fluorescence were generated for each sample. Statistical analysis was performed on the geometric scale.

Results

The equine CEA gene family cluster

The complete equine CEA gene family cluster was identified as described in "materials and methods". Surprisingly, only genes orthologous to *CEACAM1*, *CEACAM16* and *CEACAM19* exist; none orthologous gene to *CEACAM18* and *CEACAM20* could be identified. On the other hand a large number (30) of CEACAM1-related leader and N domain exons evolved in the horse genome. Further analysis indicated that 17 CEACAM1-related genes appear to be functional (named *CEACAM1*, *CEACAM41-56*) and 13 of the CEACAM1-related genes appear to be pseudogenes (named *CEACAMps1-13*) (data not shown). Based on the most recent assembly of the horse genome (EquCab2) these genes are split

into two loci by the insertion of a chromosomal region flanked by *LIPE* and *CD79A*. Similar numbers of *CEACAM1*-related genes were only found in species which contain *PSG* genes. This finding prompted us to further analyze the possible structure of the CEACAM1-related CEACAMs of the horse.

Multiple equine CEACAMs do not contain a transmembrane region

It is assumed that the main structural difference of CEACAMs concerning their function is based on the mode of their membrane anchorage. Already known CEACAMs are either type I transmembrane proteins, GPI-anchored or secreted proteins. With very few exceptions secreted CEACAMs are PSGs.

To predict the structure of equine CEACAMs we identified exons of *CEACAM1*-related genes and determined their open reading frames. Eight *CEACAM* genes contain an exon which codes for a transmembrane domain (TM exon), two of the TM exons contain a stop codon, which shortens the cytoplasmic part of the encoded domain and one has a mutation leading to a disabled splice acceptor side, indicating that this exon is not part of the processed mRNA (Fig. 2). None of the pseudogenes contains a TM exon (data not shown). In one gene (CEACAM41) the TM exon is followed by 4 exons together encoding a cytoplasmic domain (Cyt exons) possessing an immunoreceptor tyrosine-based activation motifs (ITAM) like motif. Four genes (*CEACAM1*, *42*, *43*, *50*) contain Cyt exons which code for one immunoreceptor tyrosine-based inhibition motif (ITIM) and one immunoreceptor tyrosine-based switch motif (ITSM). In *CEACAM42*, however, the first Cyt exon has no functional splice donor site. No indications were found for the existence of GPI-anchored equine CEACAMs. Remakably, seven genes (*CEACAM44*, *46*, *47*, *48*, *49*, *52*, *55*) lack a transmembrane domain encoding exon and are composed of a leader, a N domain and an IgC domain (A2-type) exon. In all these genes the third exon contains a stop codon. Since all equine CEACAMs contain a leader exon the gene products without a transmembrane domain are assumed to be secreted proteins (Fig. 2).

Phylogeny of CEACAMs in the horse

Phylogenetic trees were constructed from nucleotide sequences of the TM-, IgV-, and IgC-exons. Two

types of TM exons were identified. One (*CEACAM41*) is connected with Cyt exons encoding an ITAM-like motif and six which were related to the TM exon typically found in *CEACAM1* genes which have ITIM/ITSM-encoding exons (Fig. 3A). We have previously reported that IgV-like domains of CEACAMs with ITAM-containing supposedly activating cytoplasmic tails are closely related to IgV-like domains of inhibitory CEACAMs forming receptor pairs (Kammerer and Zimmermann 2010). Again in the present study we found that the IgV-like domain of the activating CEACAM41 is most similar to the IgV-like domains of the inhibitory receptors (CEACAM1, and CEACAM43) (Fig. 3B). Surprisingly, the IgV-like domains of secreted CEACAMs did not form one group as previously recognized for the IgV-like domains of PSGs in humans and rodents. In the horse two different groups of N domains of the secreted CEACAMs could be identified. The first is a sister group to the membrane bound CEACAMs and the second, including CEACAM47, CEACAM48 and CEACAM49, forms a group separated from all other CEACAM1-related CEACAMs. These findings may indicate that two different groups of secreted CEACAMs exist which may have different functions (Fig. 3B). The arrangement of the IgC-like domains of membrane bound CEACAMs is as in CEACAM1, starting with an A1 type domain followed by a B domain and an A2-type IgC-like domain. Interestingly, all secreted CEACAMs contain an A2-type IgC-like domain exon, which comprises a stop-codon and the 3'-untranslated region.

Generation of secreted CEACAMs in the horse

Secreted CEACAMs in the horse belong to the CEACAM1-related CEACAMs as previously found in humans and rodents. They are, however, not found in all mammalian species (Kammerer & Zimmermann 2010). During evolution, secreted CEACAMs could have been generated by duplication of membrane anchored CEACAM1. Different truncations may account for the loss of membrane anchorage of these CEACAMs in different species. Based on genomic data, generation of secreted CEACAMs in the horse was caused either by the introduction of one or more stop codons into the A2 exons or by mutation of a splice donor site of A2 domain exons. This interpretation was verified by sequencing cloned cDNAs generated from equine mRNAs (Fig. 4). In order to get experimental support for the prediction that CEACAMs without membrane anchorage are secreted, we selected four of them

and generated plasmids which encode CEACAM-GFP fusion proteins and transfected HEK-293 cells with these plasmids. As shown in Fig. 5, fusion proteins were detected in the cytoplasm of the cells in a diffuse staining pattern (Fig. 5B). Fusion proteins could further be detected in the culture supernatant of transfected cells and the amount of fusion proteins correlated with the time of culture consistent with these proteins being actively secreted by transfected cells (Fig. 5C).

Secreted CEACAMs are expressed by trophoblast cells

Next we identified tissues in which secreted equine CEACAMs are expressed. We generated gene specific primers for CEACAM44, 46, 47, 48, 49, 52, 55 and analyzed their expression by RT-PCR in 29 different equine tissues (see "Materials and Methods"). PSG-like CEACAMs were detected in a restricted number of tissues including, Endometrial cups, freshly isolated and IL-2-stimulated PBMC, lymph node, salivary gland, tongue and ovary (Fig. 6). CEACAM46 was only detected in peripheral blood mononuclear cells and CEACAM52 was not detected in any of the tissues tested. Five secreted CEACAMs (CEACAM44, 47, 48, 49, 55) were expressed in trophoblast cells or placental tissues at one or more stages of pregnancy with expression patterns differing between CEACAMs both spatially and temporally (Fig. 7). CEACAM44 was strongly expressed in day 34 conceptus tissues (chorionic girdle, allantochorion and chorion) as well as day 31, 32 and 34 cultured chorionic girdle trophoblast (all passage 0) and endometrium. CEACAM47 was only weakly expressed in day 32 and day 34 freshly isolated and cultured trophoblast cells but not in day 34 conceptus tissues or day 65 endometrial cups. CEACAM48 was weakly expressed in day 32 cultured trophoblast and day 65 endometrial cups but again not in any of the day 34 conceptus tissues. CEACAM49 was strongly expressed in day 34 chorionic girdle and cultured day 34 chorionic girdle trophoblast cells (both passage 0 and 3) and expression was maintained in day 65 endometrial cups. CEACAM55 was only expressed in day 65 endometrial cup and at a low level in PBMC of some horses. Although day 65 endometrial cup tissue may be infiltrated with lymphocytes, it is plausible that CEACAM55 expression detected in the endometrial cup tissue was due to the trophoblast cells as opposed to the expression by infiltrating lymphocytes since CEACAM46 which is preferentially expressed by leucocytes was not detected in endometrial cups (Fig 7. And data not shown). Together, these results provide supporting data that 5/7

secreted CEACAMs are PSGs.

To determine whether CEACAM44 and/or CEACAM49 expression was enriched in trophoblast cells, we then quantified their expression using qRT-PCR in matched day 34 conceptus tissue (n=3), cultured purified chorionic girdle cells (n=5), endometrium (n=3) and PBMCs (n=3). Highest expression of CEACAM44 was noted in the purified cultured passage 0 day 32 and 34 chorionic girdle trophoblast cells and day 34 chorionic girdle tissue (Fig.7). CEACAM49 expression was very specific to trophoblast tissue or cultured trophoblast cells. No significant expression was found in endometrial tissue or PBMC. In contrast to CEACAM44, CEACAM49 was expressed to a similar amount in all three different trophoblast tissues (chorionic girdle, chorion and allantochorion).

Horse PSG-like CEACAM49 has similar anti-platelet activity to human PSG1

CEACAM49 turned out to be expressed in both the chorionic girdle cells and the endometrial cups, indication that this PSG-like CEACAM may have a pivotal function. To assay the functionality CEACAM49 was subcloned from pRcCMV-ecaCEACAM49 into the pTT3 expression vector incorporating a C-terminal V5-6xHis tag (Shanley *et al.* 2013). This allowed the generation and purification of CEACAM49 protein by transient transfection of the human HEK cell line using the Freestyle™293 expression system. Eluted and dialyzed aliquots of CEACAM49 from a nickel affinity chromatography column were combined and further purified by size exclusion chromatography to remove unidentified high molecular weight bands and contaminating TGFß1 (Ballesteros *et al.* 2015) (Fig. 8A-C). To determine whether CEACAM49 exhibits similar functionality to human PSG1, the CEACAM49 protein was applied to a fluorescent fibrinogen-platelet interaction assay as described (Shanley *et al.* 2013). CEACAM49 inhibited fibrinogen binding to TRAP-activated platelets in a dose-dependent manner similar to human PSG1 (Fig. 8D, E).

Discussion

In this study, we have characterized the *CEA* gene family of the horse. Unlike all other mammals

previously investigated, we found no evidence of conserved orthologs of human *CEACAM18* and *CEACAM20* in the horse genome (Kammerer & Zimmermann 2010), consistent with the findings of a recently published independent study (Pavlopoulou & Scorilas 2014). Nevertheless, other *CEACAM1*-related *CEACAM* genes in the horse were identified and expanded significantly to probably 17 functional genes and 13 pseudogenes. Such a high number of *CEACAMs* has previously been found only in species which have an expanded *PSG* gene cluster (Kammerer & Zimmermann 2010). However, no orthologous relationship based on synteny or common structure could be identified for the horse *PSG*-like *CEACAMs*, indicating their independent evolution from a horse *CEACAM1*-like ancestor rather than a common *PSG*-like ancestor. The horse PSG-like CEACAMs are very different from PSGs in other species being composed solely of a single IgV-like N domain. Furthermore, there is significant variation in the predicted number of glycosylation sites between the equine family members, much more than is evident in the N or N1 domains of primate and rodent PSGs, respectively. This variation could reflect a diversification in function among horse PSG family members.

Until now PSGs have only been identified in mammals which have a hemochorial placenta e.g., higher primates and rodents. Hemochorial placentas are characterized by direct contact of fetal cells with the maternal blood and immune system. Therefore, potent tolerance mechanisms are required to prevent destruction of fetal cells. Since PSGs have the capacity to modulate maternal immune responses (Martinez *et al.* 2012, Martinez *et al.* 2013, Falcon *et al.* 2014) we previously suggested that maintenance of immune tolerance is an important selective force in the expansion of *PSG* gene clusters (Kammerer & Zimmermann 2010). The horse has, similar to swine and cattle, an epitheliochorial placenta in which fetal cells are separated from the maternal immune system. In addition, the recruitment of maternal immune cells into epitheliochorial placentas is very limited compared to hemochorial placentas (Carter & Enders 2013). However, the horse is exceptional in that in this species, the specialized chorionic girdle cells of the conceptus invade into the endometrium of the mare to form the endometrial cups. The dndometrial cups are recognized by maternal T-cells due to their expression of MHC class I antigens. A large number of CD4+ and CD8+ T-cells infiltrate the endometrial stroma at the edge of endometrial cups (de Mestre *et al.* 2010), and by day 60–70, these cells move into the cup.

Over the next 30 days or so, neutrophils, macrophages, and eosinophils also invade into the main area of the cup. The cups become necrotic and are eventually sloughed between days 100 and 140 in most mares. Although the immune cells are present, there is evidence that they do not attack the cup cells, and the mechanism of the demise of the cups is currently unclear. However, despite such a close contact with maternal immune cells, semiallogeneic fetal cells are not attacked, indicating that powerful tolerance mechanisms are at work (Noronha & Antczak 2010).

Several immune tolerance mechanisms during pregnancy were recently identified in primates and rodents, including downregulation of classical MHC class I antigens, upregulation of indoleamine 2,3-dioxygenase (IDO) which causes a local depletion of tryptophan, and upregulation of CD95L (Fas ligand) that promotes apoptosis of activated lymphocytes (Trowsdale & Betz 2006). Additional poorly characterized secreted factors are likely to participate in tolerance induction during pregnancy (Trowsdale & Betz 2006). PSGs may represent such factors inducing immunotolerance and thromboregulation, in human and mice (Moore & Dveksler 2014, Ballesteros *et al.* 2015). Thus, we speculate that equine PSG-like CEACAMs expressed by infiltrating chorionic girdle trophoblast and endometrial cup cells may have a similar function controlling activity of T cells located at the borders of endometrial cups, and perhaps on peripheral T cells. Indeed, peripheral blood lymphocytes from pregnant mares have a reduced capacity to develop into effective cytotoxic T cells (Baker *et al.* 1999). PSGs have the potential to be present in maternal blood, as one of the major secretions of endometrial cups is equine chorionic gonadotropin, which is found at very high levels in the blood of pregnant mares while the cups persist (Allen 2001). Although additional work is needed to verify putative immunoregulatory functions of PSGs, in this study, we did confirm that PSG-like equine CEACAM49, similar to human and mouse PSGs, inhibits the interaction between activated platelets and fibrinogen (Shanley *et al.* 2013). An anti-platelet activity may be beneficial for endometrial cup function because at the base of endometrial cups there are many newly formed lymph and blood vessels to support constant circulation in this immature capillary bed (Antczak *et al.* 2013). In addition, PSGs have been shown to bind other integrins, besides platelet integrin (Moore & Dveksler 2014), and it is possible that this also occurs in the horse. Therefore, a PSG-integrin interaction may have additional yet unknown

benefits for the function and/or survival of endometrial cups. Thus equine PSGs have important functional similarities with human and mouse PSGs indicating that they were generated independently by convergent evolution.

There is probably no recent common PSG ancestor from which the PSGs in different species evolved. Lack of synteny suggests that PSGs evolved independently in different mammalian species. As a consequence PSGs structurally differ between primates, rodents and perissodactyls. Nevertheless, molecular interaction of PSGs with integrins is found in all mentioned orders. This striking similarity may be explained by an interaction via the N-terminal IgV-like domain which is present in all types of PSGs. Interaction with integrins could be part of the core functional requirements for the selection of PSGs gene expansion. Whether regulation of immunity belongs also to these core functions, has to be confirmed in further studies. Remarkably, the horse is an exceptionally well suited animal model to address such a question, since mares mount a strong systemic (antibody) and local (cellular) immune response to paternal antigens (Noronha & Antczak 2010). This naturally occurring immune recognition will be beneficial for the determination of the immune regulatory function of PSGs. Whatever the core functions of PSGs are, there must be an extremely strong selection pressure to select for the expansion of the *PSG* cluster in disparate species either for diversification and functional fine tuning or for increase of the dose of highly similar proteins. Among the 12 protein-coding genes with the highest level of enriched expression in human placenta compared to other tissues are eight *PSG* genes (Uhlen *et al.* 2015). A better understanding of PSG function in horses may have important implications for the elucidation of the role of PSGs in human pregnancy.

Availability of supporting data

Sequences of cloned equine CEACAMs were submitted to GenBank and have the following accession numbers: Eca_CEACAM46a, KT124640; Eca_CEACAM46b, KT124641; Eca_CEACAM47 KT124642; Eca_CEACAM49, KT124643; Eca_CEACAM55, KT124644.

Competing interests

The author(s) declare that they have no competing interests.

Abbreviations

CEA, carcinoembryonic antigen; CEACAM, carcinoembryonic antigen-related cell adhesion molecule; Cyt, cytoplasmic domain; GPI, glycosylphosphatidyl inositol; ITAM, immunoreceptor tyrosine-based activation motif; ITIM, immunoreceptor tyrosine-based inhibition motif; ITSM, immunoreceptor tyrosine-based switch motif; PSG, pregnancy-specific glycoprotein; TM, transmembrane domain.

Authors' contributions

DA and LB did most of the experiments. SM, JH, WW and JR did additional experiments. JH and VCS collected tissues and JH performed the biopsy of the endometrial cups. RO'R conducted platelet studies. WZ,TM and AdM analyzed data and participated in manuscript writing. RK conceived the study, did biocomputing and manuscript writing. All authors participated in the design of the study, and read and approved the final version.

Acknowledgements

We thank Andrea Braun for expert technical assistance and Prof. Niamh Moran for facilitating platelet experiments. TM acknowledges funding from an Irish Higher Education Authority PRTLI3 (HEA-PRTLI3) grant and an Investigator Award from Science Foundation Ireland. AdM acknowledges funding from the Wellcome Trust (WT09158). The study was further supported by GIZ (Contract no. 81170269; Project No. 13.1432.7---001.00) and by BMWi (ZIM) KF2875802UL2 to RK.

References

Allen WR 2001 Fetomaternal interactions and influences during equine pregnancy. *Reproduction* **121** 513-527.

Allen WR, Hamilton DW & Moor RM 1973 The origin of equine endometrial cups. II. Invasion of the endometrium by trophoblast. *Anat Rec* **177** 485-501.

Antczak DF, de Mestre AM, Wilsher S & Allen WR 2013 The equine endometrial cup reaction: a fetomaternal signal of significance. *Annu Rev Anim Biosci* **1** 419-442.

Arnold LL, Doherty TM, Flor AW, Simon JA, Chou JY, Chan WY & Mansfield BC 1999 Pregnancy-specific glycoprotein gene expression in recurrent aborters: a potential correlation to interleukin-10 expression. *Am J Reprod Immunol* **41** 174-182.

Baker JM, Bamford AI & Antczak DF 1999 Modulation of allospecific CTL responses during pregnancy in equids: an immunological barrier to interspecies matings? *J Immunol* **162** 4496-4501.

Ballesteros A, Mentink-Kane MM, Warren J, Kaplan GG & Dveksler GS 2015 Induction and activation of latent transforming growth factor-beta1 are carried out by two distinct domains of pregnancy-specific glycoprotein 1 (PSG1). *J Biol Chem* **290** 4422-4431.

Bebo BF, Jr. & Dveksler GS 2005 Evidence that pregnancy specific glycoproteins regulate T-Cell function and inflammatory autoimmune disease during pregnancy. *Curr Drug Targets Inflamm Allergy* **4** 231-237.

Bohn H 1971 [Detection and characterization of pregnancy proteins in the human placenta and their quantitative immunochemical determination in sera from pregnant women]. *Arch Gynakol* **210** 440-457.

Cabrera-Sharp V, Read JE, Richardson S, Kowalski AA, Antczak DF, Cartwright JE, Mukherjee A & de Mestre AM 2014 SMAD1/5 signaling in the early equine placenta regulates trophoblast differentiation and chorionic gonadotropin secretion. *Endocrinology* **155** 3054-3064.

Camolotto S, Racca A, Rena V, Nores R, Patrito LC, Genti-Raimondi S & Panzetta-Dutari GM 2010 Expression and transcriptional regulation of individual pregnancy-specific glycoprotein genes in differentiating trophoblast cells. *Placenta* **31** 312-319.

Carter AM & Enders AC 2013 The evolution of epitheliochorial placentation. *Annu Rev Anim Biosci* **1** 443-467.

Chuong EB, Tong W & Hoekstra HE 2010 Maternal-fetal conflict: rapidly evolving proteins in the rodent placenta. *Mol Biol Evol* **27** 1221-1225.

de Mestre A, Noronha L, Wagner B & Antczak DF 2010 Split immunological tolerance to trophoblast. *Int J Dev Biol* **54** 445-455.

de Mestre AM, Bacon SJ, Costa CC, Leadbeater JC, Noronha LE, Stewart F & Antczak DF 2008 Modeling trophoblast differentiation using equine chorionic girdle vesicles. *Placenta* **29** 158-169.

Donaldson WL, Zhang CH, Oriol JG & Antczak DF 1990 Invasive equine trophoblast expresses conventional class I major histocompatibility complex antigens. *Development* **110** 63-71.

Falcon CR, Martinez FF, Carranza F, Cervi L & Motran CC 2014 In vivo expression of recombinant pregnancy-specific glycoprotein 1a inhibits the symptoms of collagen-induced arthritis. *Am J Reprod Immunol* **72** 527-533.

Fialova L, Kohoutova B, Peliskova Z, Malbohan I & Mikulikova L 1991 [Serum levels of trophoblast-specific beta-1-globulin (SP1) and alpha-1-fetoprotein (AFP) in pregnant women with rheumatoid arthritis]. *Cesk Gynekol* **56** 166-170.

Hemberger M 2013 Immune balance at the foeto-maternal interface as the fulcrum of reproductive success. *J Reprod Immunol* **97** 36-42.

Hubert MA, Sherritt SL, Bachurski CJ & Handwerger S 2010 Involvement of transcription factor NR2F2 in human trophoblast differentiation. *PLoS One* **5** e9417.

Kammerer R, Popp T, Hartle S, Singer BB & Zimmermann W 2007 Species-specific evolution of immune receptor tyrosine based activation motif-containing CEACAM1-related immune receptors in the dog. *BMC Evol Biol* **7** 196.

Kammerer R, Popp T, Singer BB, Schlender J & Zimmermann W 2004 Identification of allelic variants of the bovine immune regulatory molecule CEACAM1 implies a pathogen-driven evolution. *Gene* **339** 99-109.

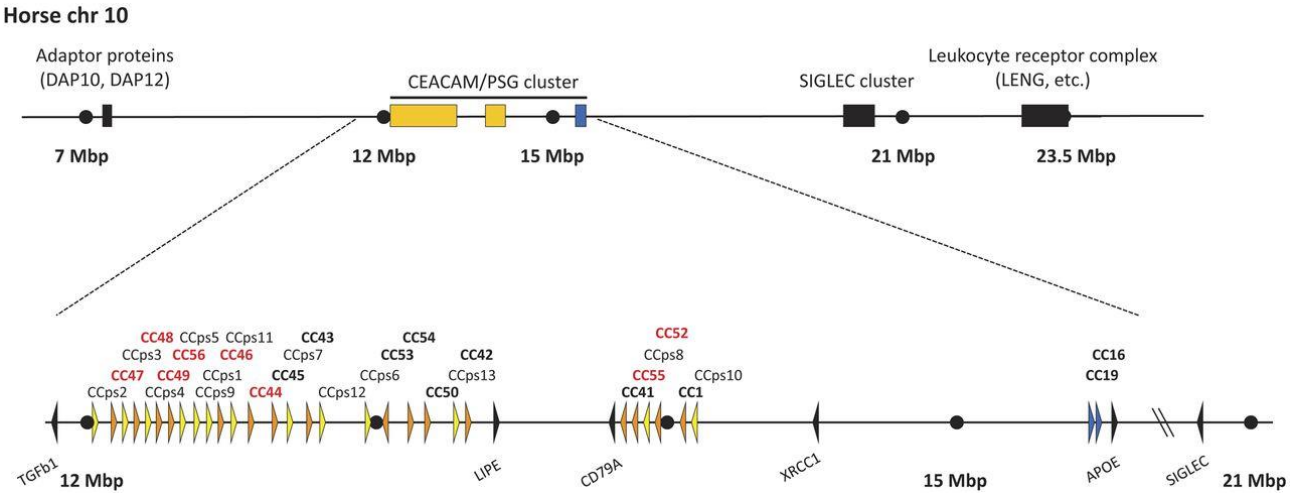
Kammerer R & Zimmermann W 2010 Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families. *BMC Biol* **8** 12.

Karg NJ, Csaba IF, Than GN, Arany AA & Szabo DG 1981 The prognosis of the possible foetal and placental complications during delivery by measuring maternal serum levels of pregnancy-

specific beta-1-glycoprotein (SP1). *Arch Gynecol* **231** 69-73.
Lin TM, Halbert SP & Spellacy WN 1974 Measurement of pregnancy-associated plasma proteins during human gestation. *J Clin Invest* **54** 576-582.
Lunn P, Vagnoni KE & Ginther OJ 1997 The equine immune response to endometrial cups. *J Reprod Immunol* **34** 203-216.
Martinez FF, Cervi L, Knubel CP, Panzetta-Dutari GM & Motran CC 2013 The role of pregnancy-specific glycoprotein 1a (PSG1a) in regulating the innate and adaptive immune response. *Am J Reprod Immunol* **69** 383-394.
Martinez FF, Knubel CP, Sanchez MC, Cervi L & Motran CC 2012 Pregnancy-specific glycoprotein 1a activates dendritic cells to provide signals for Th17-, Th2-, and Treg-cell polarization. *Eur J Immunol* **42** 1573-1584.
McLellan AS, Fischer B, Dveksler G, Hori T, Wynne F, Ball M, Okumura K, Moore T & Zimmermann W 2005 Structure and evolution of the mouse pregnancy-specific glycoprotein (Psg) gene locus. *BMC Genomics* **6** 4.
Moffett A & Loke C 2006 Immunology of placentation in eutherian mammals. *Nat Rev Immunol* **6** 584-594.
Moore T & Dveksler GS 2014 Pregnancy-specific glycoproteins: complex gene families regulating maternal-fetal interactions. *Int J Dev Biol* **58** 273-280.
Motran CC, Diaz FL, Montes CL, Bocco JL & Gruppi A 2003 In vivo expression of recombinant pregnancy-specific glycoprotein 1a induces alternative activation of monocytes and enhances Th2-type immune response. *Eur J Immunol* **33** 3007-3016.
Noronha LE & Antczak DF 2010 Maternal immune responses to trophoblast: the contribution of the horse to pregnancy immunology. *Am J Reprod Immunol* **64** 231-244.
Pavlopoulou A & Scorilas A 2014 A comprehensive phylogenetic and structural analysis of the carcinoembryonic antigen (CEA) gene family. *Genome Biol Evol* **6** 1314-1326.
Rawn SM & Cross JC 2008 The evolution, regulation, and function of placenta-specific genes. *Annu Rev Cell Dev Biol* **24** 159-181.
Schumacher A, Costa SD & Zenclussen AC 2014 Endocrine factors modulating immune responses in pregnancy. *Front Immunol* **5** 196.
Shanley DK, Kiely PA, Golla K, Allen S, Martin K, O'Riordan RT, Ball M, Aplin JD, Singer BB, Caplice N, Moran N & Moore T 2013 Pregnancy-specific glycoproteins bind integrin alphaIIbbeta3 and inhibit the platelet-fibrinogen interaction. *PLoS One* **8** e57491.
Silver RM, Heyborne KD & Leslie KK 1993 Pregnancy specific beta 1 glycoprotein (SP-1) in maternal serum and amniotic fluid; pre-eclampsia, small for gestational age fetus and fetal distress. *Placenta* **14** 583-589.
Snyder SK, Wessner DH, Wessells JL, Waterhouse RM, Wahl LM, Zimmermann W & Dveksler GS 2001 Pregnancy-specific glycoproteins function as immunomodulators by inducing secretion of IL-10, IL-6 and TGF-beta1 by human monocytes. *Am J Reprod Immunol* **45** 205-216.
Towler CM, Horne CH, Jandial V, Campbell DM & MacGillivray I 1977 Plasma levels of pregnancy-specific beta 1-glycoprotein in complicated pregnancies. *Br J Obstet Gynaecol* **84** 258-263.
Trowsdale J & Betz AG 2006 Mother's little helpers: mechanisms of maternal-fetal tolerance. *Nat Immunol* **7** 241-246.
Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J & Ponten F 2015 Proteomics. Tissue-based map of the human proteome. *Science* **347** 1260419.
von Rango U 2008 Fetal tolerance in human pregnancy--a crucial balance between acceptance and limitation of trophoblast invasion. *Immunol Lett* **115** 21-32.
Wessells J, Wessner D, Parsells R, White K, Finkenzeller D, Zimmermann W & Dveksler G 2000 Pregnancy specific glycoprotein 18 induces IL-10 expression in murine macrophages. *Eur J Immunol* **30** 1830-1840.
Zebhauser R, Kammerer R, Eisenried A, McLellan A, Moore T & Zimmermann W 2005 Identification of a novel group of evolutionarily conserved members within the rapidly diverging murine Cea family. *Genomics* **86** 566-580.

Figure Legends

Figure 1 - Genomic arrangement of the extended Leukocyte Receptor Complex and the CEACAM/PSG locus in the horse.



The CEACAM/PSG cluster is located on horse chromosome 10 within the extended leukocyte receptor complex (LRC). The extended LRC is limited on both ends by the genes coding for adaptor proteins and the Leukocyte receptor complex, respectively. Loci containing genes of the CEACAM1-related CEACAMs are shown in yellow and the locus containing the conserved CEACAM genes are shown in blue. Lower part: Genomic organization of horse CEA gene family loci. Arrowheads represent genes with their transcriptional orientation. CEACAM pseudogenes are shown in yellow and named in black. Putatively intact CEACAM genes are shown in orange and named in bold (secreted CEACAMs in red and membrane anchored in black). SIGLEC genes and marker genes are shown as black arrowheads. The scale indicated by dots is 1 Mbp unless interrupted by slanted lines. The horse Ensembl/NCBI release EquCab2 was used. Nucleotide numbering of the chromosomes starts at the centromere. CCps, CEACAM pseudogenes; CC, CEACAM; chr, chromosome.

Figure 2 – Prediction of CEA gene family members from sequence analyses of the horse genome. Exon arrangement of equine CEACAM1-related genes. The exon types are indicated by differently colored boxes. Leader exons are shown as gray, IgV-like domain exons as red, IgC-like domain exons as blue and transmembrane domain exons as black boxes. The exons encoding the cytoplasmic domain with an ITIM or an ITAM motif are shown in yellow and green, respectively. The presence of nonsense mutations, deletions/insertions in exons causing reading frame shifts are indicated by asterisks. Two genes with intact ITIM/ITSM were identified and named CEACAM1 and CEACAM43. No leader exon could be identified for CEACAM51 because of a sequence gap in the publicly available genomic sequences. The genes are arranged in the order and orientation as found on horse chromosome 10. C1-C4, cytoplasmic domain exons; L, leader exon; TM, transmembrane exon; ~ mutation of splice sites.

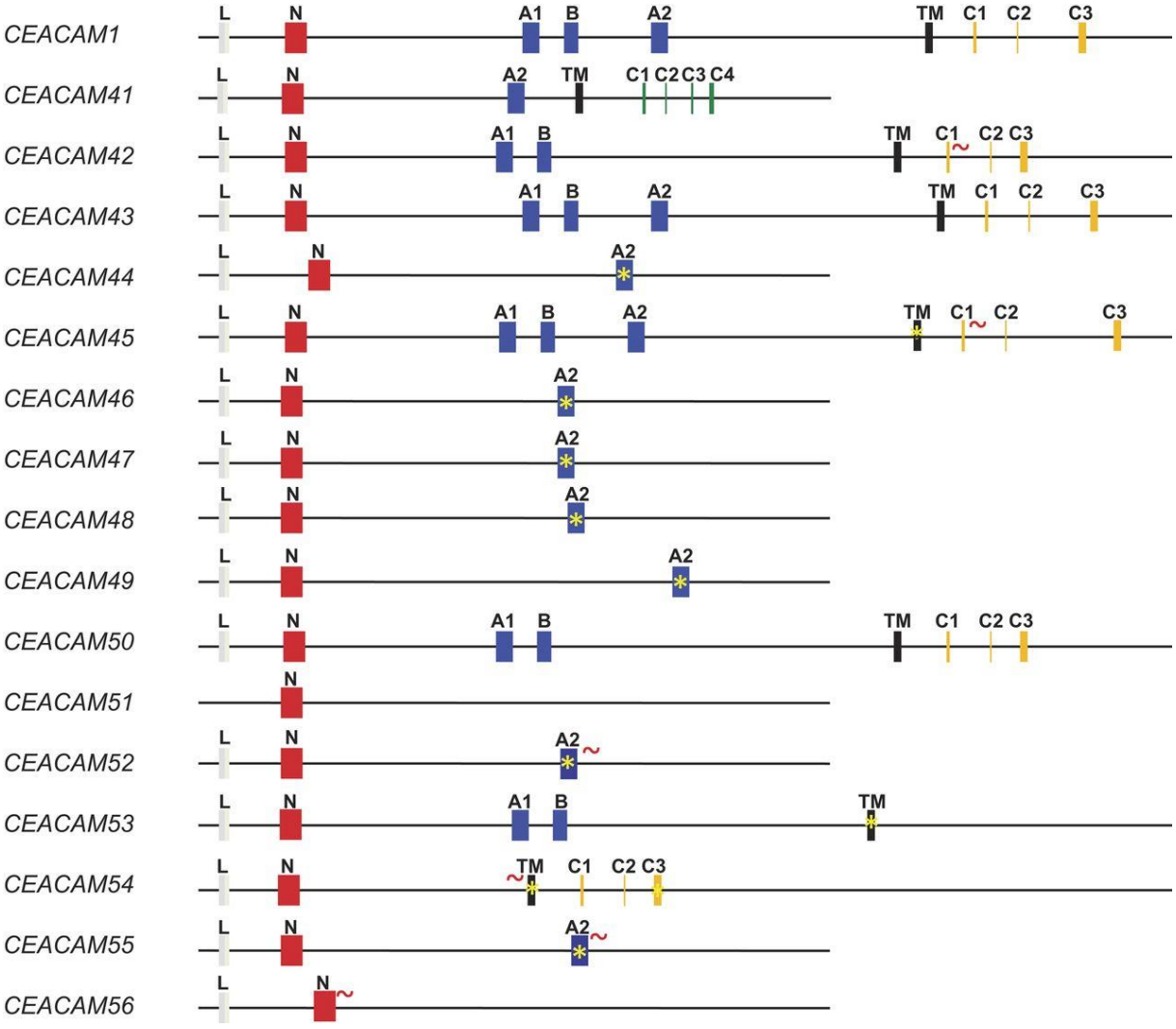
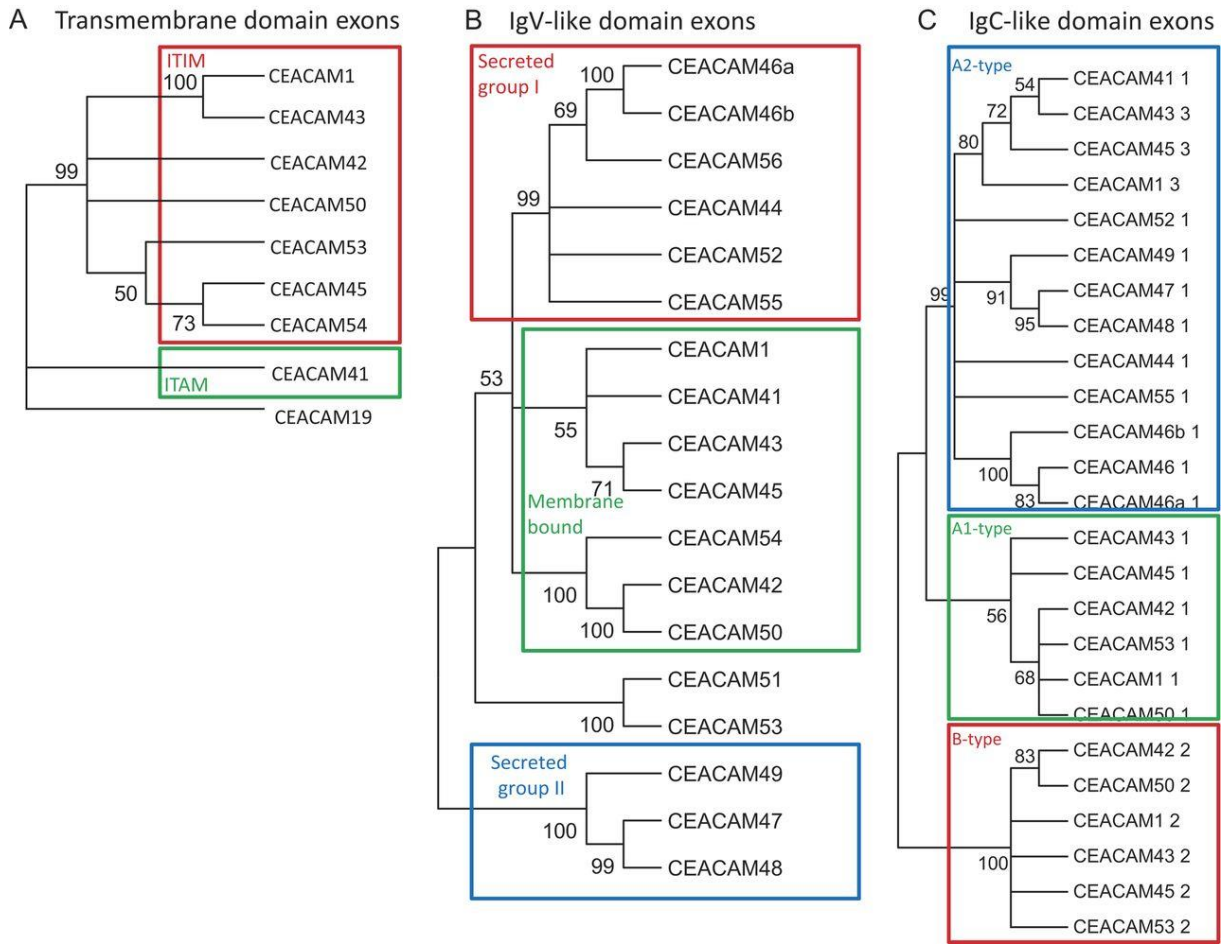


Figure 3 - Phylogeny of equine CEACAMs.

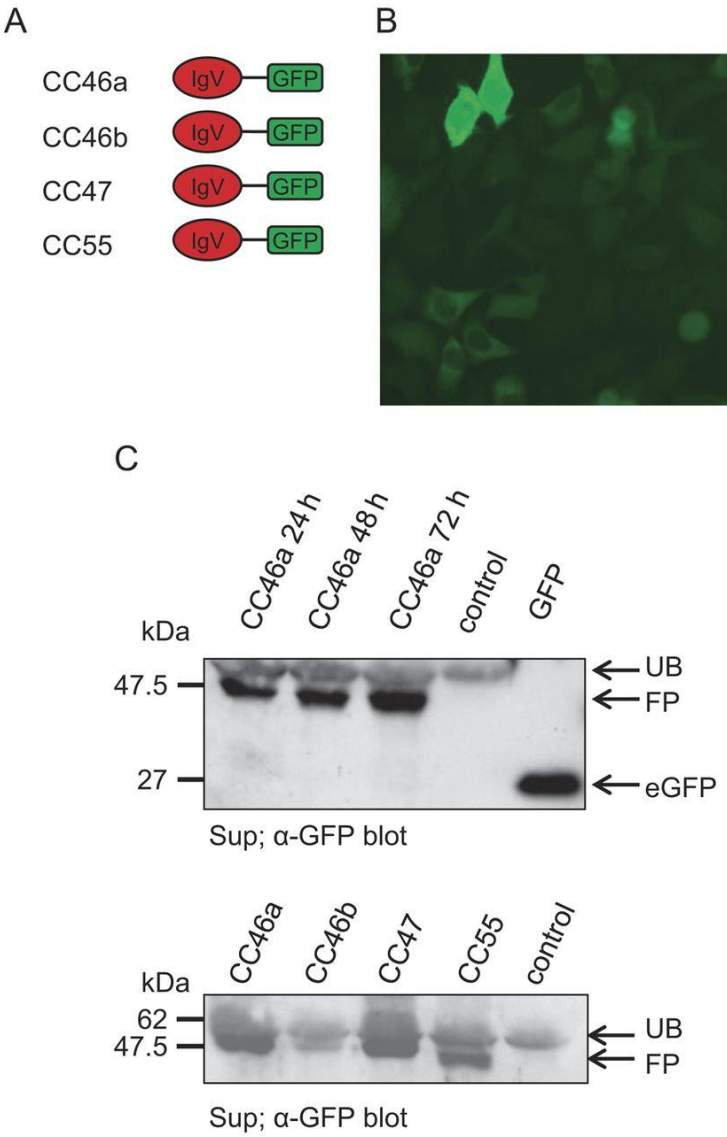


Phylogenetic trees were constructed from transmembrane domain exon (A), IgV-like domain exon (B) and IgC-like domain exon (C) nucleotide sequences from CEACAM members of the horse using the UPGMA method (MEGA 6.0 software). The reliability of a phylogenetic tree was assessed using the Bootstrap test applying 500 replicates. The statistical support for selected nodes is shown. Bootstrap values >50 are shown. CEACAM46a and CEACAM46b are putative allelic CEACAM46 variants. Boxes group CEACAMs with the indicated properties.

Figure 4 – Mechanisms for generation of secreted CEA family members in the horse.

Secreted CEACAMs in the horse are formed by two different mechanisms: firstly by inclusion of A2-type domain exons (exon 3), which contain multiple stop codons and secondly by mutation of the splice donor site at the end of the N domain exon, leading to read through into the following intron. CEACAMs from which the structure was verified by cloning are underlined. Stop codons are shown in red. Homologous codons of other CEACAMs representing non-stop codons are indicated in blue. The mutated splice donor site of CEACAM56 is indicated in green.

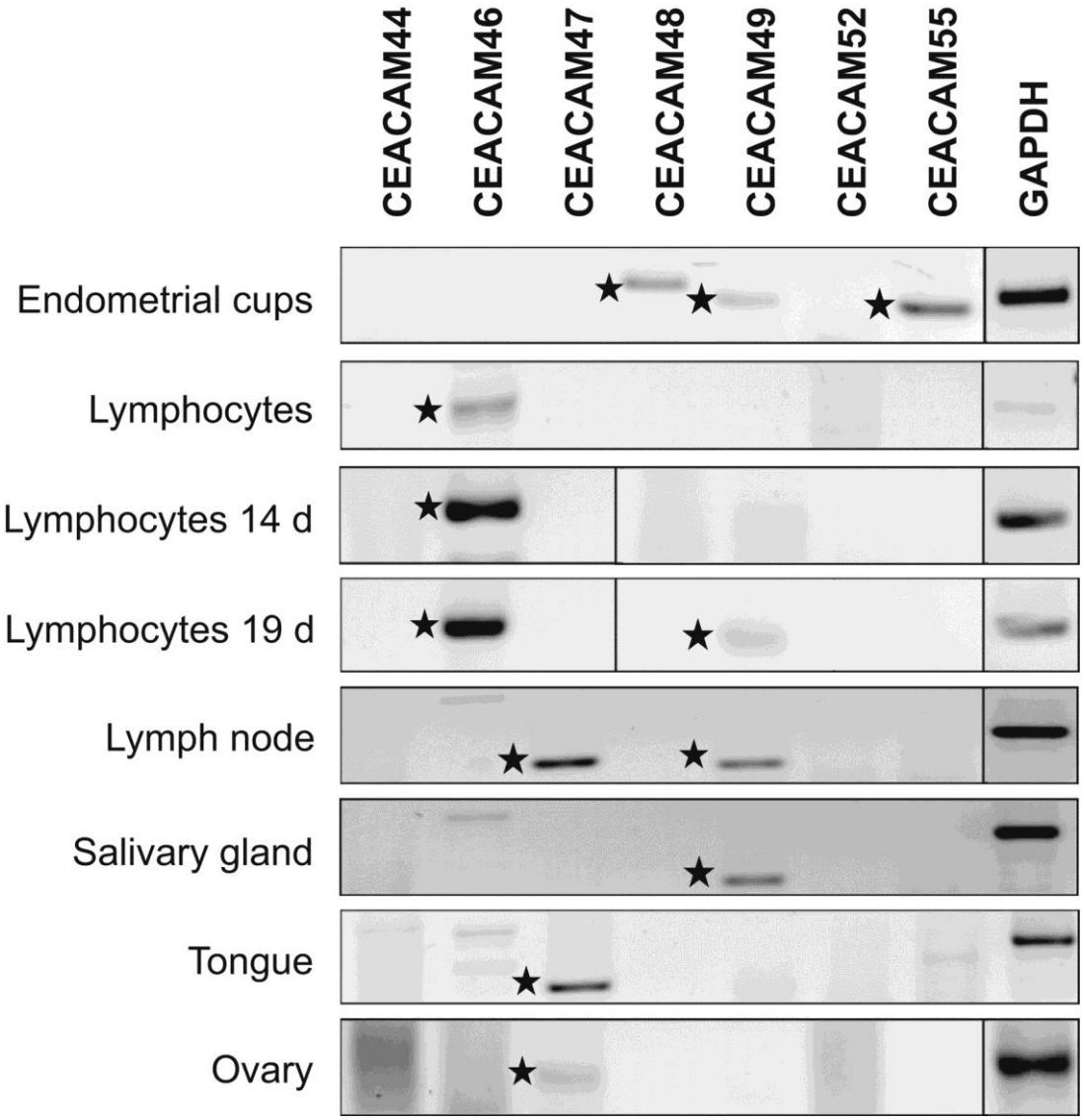
Figure 5: PSG-like CEACAMs are secreted by eukaryotic cells.



(A) Schematic illustration of the proteins used in the study. CEACAM46a, CEACAM46b, CEACAM47, and CEACAM55 were expressed in human cells as GFP fusion proteins. (B) Expression of CEACAM46a-eGFP fusion protein by transfected and sorted HEK-293A cells. GFP fusion proteins are only present in the cytoplasm not in the nucleus as wild type GFP or at the cell surface as membrane anchored CEACAMs. (C) Western blot analysis to detect secreted CEACAM-eGFP fusion proteins in the culture supernatant of transfected HEK-293A cells. Supernatant of untransfected HEK-293A cells was used as control. Upper panel: the amount of fusion proteins in the culture supernatant increased with time of culture. For positive control of eGFP detection cell extracts of eGFP transfected HEK-293A cells were used. Lower panel: All constructs accumulate in the culture medium. CC46a, CEACAM46_eGFP; CC46b, CEACAM46b_eGFP; CC47, CEACAM47_eGFP; CC55,

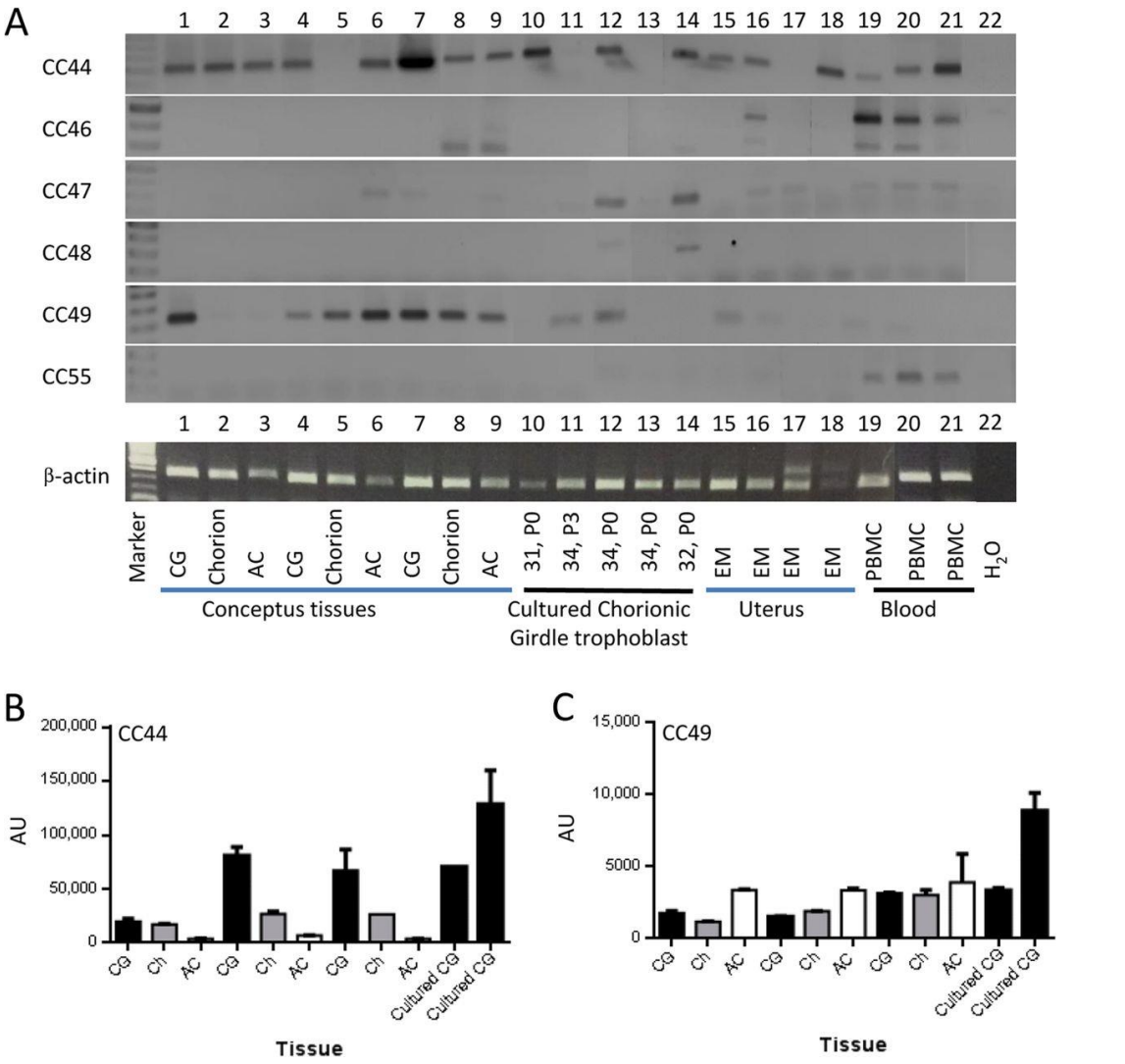
CEACAM55_eGFP; Sup, supernatant; UB, unspecific binding; FP, fusion protein;

Figure 6 - Horse PSG-like CEACAMs are preferentially expressed in PBMC and endometrial cups.



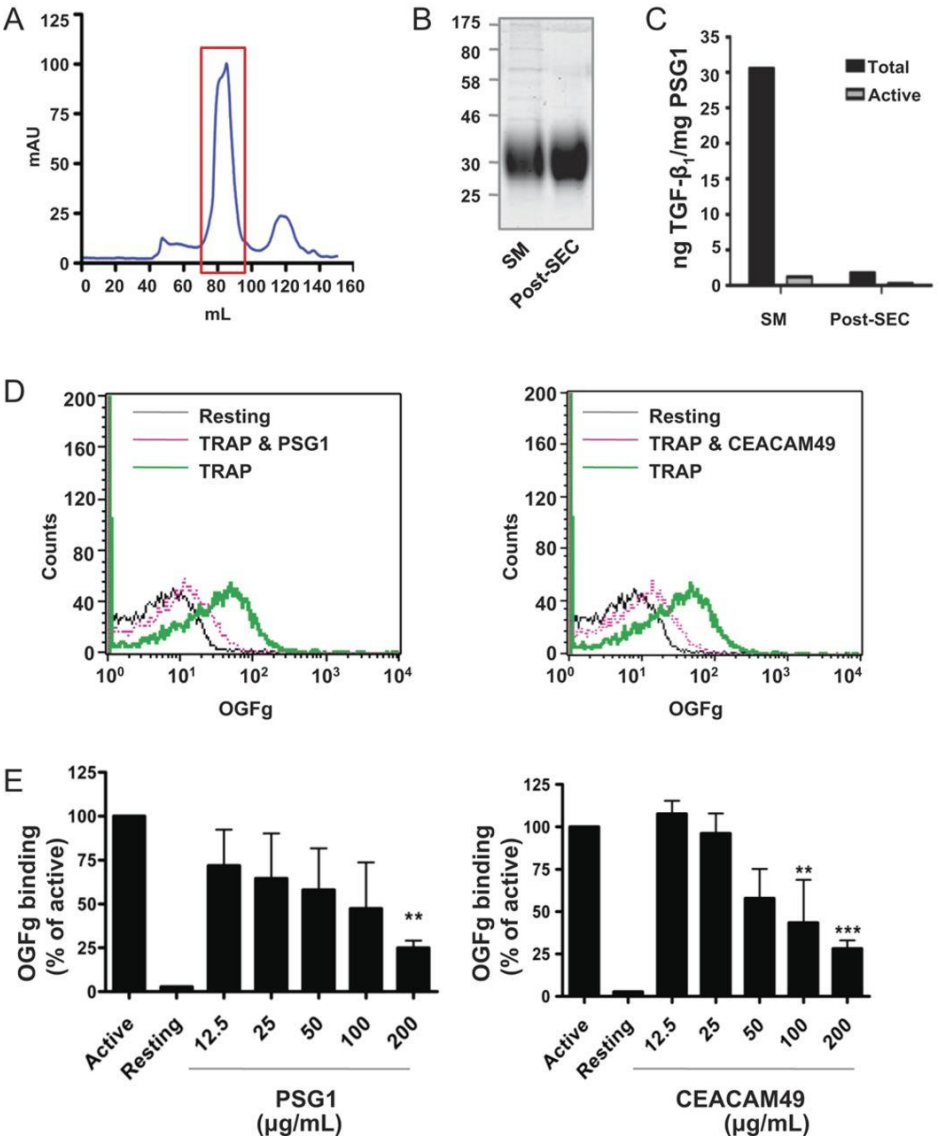
Total RNA was isolated from various equine tissues as described in "Material and Methods". Biopsies of endometrial cups were taken on day 65 of pregnancy. Lymphocytes were isolated from peripheral blood of healthy horses some of these lymphocytes were cultured in the presence of recombinant human IL-2 for 14 or 19 days (d). In additional 25 equine tissues ("Materials and Methods") none of these mRNAs were detected. Gene-specific primers used in these experiments are summarized in Table 1. Asterix indicate bands of the expected size.

Figure 7 – Horse PSG-like molecules are expressed in trophoblast cells.



A. Qualitative RT-PCR analysis of expression of secreted CEACAMs in day 34 conceptus tissues (n=3), cultured day 31-34 chorionic girdle trophoblast cells either isolated without prior passaging (P0) or isolated after 3 passages (P3) (n=5), non-pregnant endometrium (n=4) and PBMC isolated from non-pregnant mares (n=3). Amplicons were generated using primers specific for equine CEACAM44, 46, 47, 48, 49, 55, and β-ACTIN (as control) mRNA, and all bands observed in all tissues with each primer set are of the correct predicted size (table 1). CG=chorionic girdle, AC=allantochorion, EM=Endometrium, H_20 = no template control. B. and C. Real time qRT-PCR expression of CEACAM44 (B) and CEACAM49 (C) mRNA in three independent day 34 conceptus tissues, chorionic girdle (CG), chorion (Ch), allantochorion (AC) as well as cultured day 32 and 34 chorionic girdle trophoblast cells (n=2). Data points represent the mean +/- the SE of two experimental replicates.

Figure 8 – Horse PSG-like CEACAM49 has similar anti-platelet activity to human PSG1.



(A) Chromatogram of equine CEACAM49 applied to a Superdex 200 prep grade size exclusion column with red box indicating fractions that were pooled and concentrated. (B) Coomassie stained 12% polyacrylamide gel with 8 µg of indicated protein per lane. (C) TGFβ1 ELISA of SM and post-SEC equine CEACAM49 preparations. (D) FACS analysis showing fluorescent Oregon Green Fibrinogen (OGFg) binding to resting and thrombin receptor activating peptide (TRAP)-activated human platelets with, and without, 200 µg/ml human PSG1 or equine CEACAM49 pre-treatment. (E) Graphs of dose-dependent inhibition of OGFg binding to activated human platelets, relative to untreated controls, by human PSG1 and equine CEACAM49. n = 3; One-way ANOVA; **, P<0.01; ***, P<0.001. SM, starting material; SEC, size exclusion chromatography.

Tables

Table 1: Gene-specific oligonucleotides for expression analyses and cDNA cloning of horse CEA gene family members

gene	oligonucleotide sequence	location of primers (exon)	size of PCR product (bp)
CEACAM1	For: TGCATCATATAAGATAGGCCCAG Rev: AGTGAGAGTCCTCTTGTCCAGG	N domain A1 domain	367
CEACAM41	For: CATTGCATGTGATAGAGCGAC Rev: CGTCCTTCTGTTCTGTGACTGT	N domain A2 domain	246
CEACAM42	For: AGGGGAAGGAATAGATCCCG Rev: GAGTCCTGTTGTCCGAGGATAG	N domain A1 domain	392
CEACAM43	For: CCCATCAAGAAATTGTGTCCT Rev: ATGTTAACACTACAGGGTCCCTG	N domain A1 domain	276
CEACAM44	For: GCTGTTGTAGGGACCGATGTTA Rev: CCTCCTTCCTGATGATGTGTGT	N domain A2 domain	503
CEACAM45	For: GCGATAGGGCAACAAGAAATTAT Rev: CATGTTAACACTACAGGGTCCTCA	N domain A1 domain	254
CEACAM46	For: AGTCCCACCCAATGGTATCC Rev: CCCAAGTATTGCCCCTTCTGT	N domain A1 domain	525
CEACAM47	For: ACAGACCAAGTCCCAAAACC Rev: TTACGTCTTTTGACGATTCCAG	N domain A1 domain	256
CEACAM48	For: GACCAGCTCGCAAACAAA Rev:	N domain A2 domain	279

	AACAAGCTTCTTATCTGGCATTT		
CEACAM49	For: TGGAGCACGTCCACATAAAC Rev: GGAGGTATTTGACCCTGGATT	N domain A2 domain	232
CEACAM50	For: AGATGCTCTTGAAGGAACGGAT Rev: GACAGCTTCAGCCAGGTCCTA	N domain A1 domain	453
CEACAM51	ND	NA	NA
CEACAM52	For: ATGCTGCTGCAGGGGATA Rev: CATCCTCCCTCCTGACACAT	N domain A2 domain	508
CEACAM53	For: TTCAAAGGGGAAATAGATTCCA Rev: GAGTCCTGTTGTCTGGGGAC	N domain A1 domain	394
CEACAM54	For: ATCAGTCCCTGGCTTCAGA Rev: TACACGGAGCTGTATACTTC	N domain A domain	409
CEACAM55	For: CCCTACTAGTCACGAGGAAGAAC Rev: CATCCTCTCGGTCAGTCACA	N domain A2 domain	222
CEACAM46a	For: TCTTCTCACAGAGGGGAGGA Rev: GGGCCAGACTCATATTTCACA	5'UTR 3'UTR	777
CEACAM46b	For: TCTTCTCACAGAGGGGAGGA Rev: GGGCCAGACTCATATTTCACA	5'UTR 3'UTR	777
CEACAM47	For: CCAGGCTCTTCTCACAGAGG Rev: CGTCTTTTGACGATTCCAGTG	5'UTR 3'UTR	588
CEACAM49	For: CCAAGCTCTTCTCACAGACG Rev: GCGGCTCTGTAACTGGGGTCTT	5'UTR 3'UTR	600
CEACAM55	For: GGGCATAAGCTCTTCTCACG Rev: GTATATGTGGAGCTCTCCAA	5'UTR 3'UTR	714

1 ND, not done; NA, not applicable; For: forward primer; Rev: reverse primer

2