molecular systems biology

Comparative systems biology of human and mouse as a tool to guide the modeling of human placental pathology

Brian Cox[1],*, Max Kotlyar[2,3], Andreas I Evangelou[4], Vladimir Ignatchenko[4], Alex Ignatchenko[4], Kathie Whiteley[5], Igor Jurisica[2,3,6], S Lee Adamson[5,7], Janet Rossant[1,8] and Thomas Kislinger[3,4],*

[1] The Hospital for Sick Children, Program in Developmental and Stem Cell Biology, Toronto, Ontario, Canada, [2] Division of Signaling Biology, Ontario Cancer Institute, Toronto, Ontario, Canada, [3] Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada, [4] Division of Cancer Genomics and Proteomics, Ontario Cancer Institute, Toronto, Ontario, Canada, [5] Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada, [6] Department of Computer Science, University of Toronto, Toronto, Ontario, Canada, [7] Department of Obstetrics and Gynaecology, University of Toronto, Toronto, Ontario, Canada and [8] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada
* Corresponding authors. B Cox, Toronto Medical Discovery Tower, SickKids Research Institute, 101 College Street, Room 13-305, Toronto, Ontario, Canada M5G 1L7. Tel.: + 416 358 5763; Fax: + 416 813 5252; E-mail: b.cox@utoronto.ca or T Kislinger, Toronto Medical Discovery Tower, Ontario Cancer Institute, 101 College Street, Room 9-807, Toronto, Ontario, Canada. Tel.: + 416 581 7627; Fax: + 416 581 7629; E-mail: thomas.kislinger@utoronto.ca

Placental abnormalities are associated with two of the most common and serious complications of human pregnancy, maternal preeclampsia (PE) and fetal intrauterine growth restriction (IUGR), each disorder affecting ~5% of all pregnancies. An important question for the use of the mouse as a model for studying human disease is the degree of functional conservation of genetic control pathways from human to mouse. The human and mouse placenta show structural similarities, but there have been no systematic attempts to assess their molecular similarities or differences. We collected protein and mRNA expression data through shot-gun proteomics and microarray expression analysis of the highly vascular exchange region, microdissected from the human and mouse near-term placenta. Over 7000 ortholog genes were detected with 70% co-expressed in both species. Close to 90% agreement was found between our human proteomic results and 1649 genes assayed by immunohistochemistry for expression in the human placenta in the Human Protein Atlas. Interestingly, over 80% of genes known to cause placental phenotypes in mouse are co-expressed in human. Several of these phenotype-associated proteins form a tight protein–protein interaction network involving 15 known and 34 novel candidate proteins also likely important in placental structure and/or function. The entire data are available as a web-accessible database to guide the informed development of mouse models to study human disease.
Molecular Systems Biology **5**: 279; published online 16 June 2009; doi:10.1038/msb.2009.37
Subject Categories: proteomics; functional genomics
Keywords: IUGR; orthologs; placenta; preeclampsia; proteomics

Introduction

An important question for the development of human disease models is the knowledge of functional conservation for candidate human disease genes in the model organism. The mouse has proven useful for studying the genetic basis of human disease, although there have been cases in which mutations in the orthologous mouse genes did not recapitulate the human disease state (Benfey and Mitchell-Olds, 2008). These differences may be due to changes in the conservation of genetic control pathways from human to mouse. Therefore, careful analyses of the expression of orthologous genes and

their interaction networks between human and mouse tissues are required to determine the full utility of the mouse as an organism to model human diseases. The only published direct comparison between human and mouse, on a proteomic level, is a recent publication of red blood cells (RBCs) by Mann and colleagues (Pasini *et al*, 2008). In this study, we chose placenta as our model of choice for several reasons (1) healthy tissue from human and mouse is readily available for in-depth proteomics analyses, as compared with most other tissues, (2) it is a complex organ comprising of many supporting cell types, including a complex vascular network and (3) poor mechanistic understandings of placental vascular diseases

argue for detailed dissection of these malfunctions in functional mouse models.

The development of a functional placenta is critical for a successful pregnancy. It is the site of feto-maternal transport of gases, nutrients and metabolites, it protects the conceptus from the immune system of the mother and it produces hormones that regulate maternal adaptations to pregnancy. Placental dysfunction is believed to be the major cause of two of the most common and serious complications of human pregnancy, maternal preeclampsia (PE) and fetal intrauterine growth restriction (IUGR). Each disorder affects ∼5% of all pregnancies and is associated with serious morbidity and mortality, and the only known treatment is premature delivery, which places the baby at high risk of prematurity-related complications (Roberts and Gammill, 2005; Mari and Hanif, 2007). Although PE and IUGR have been intensely studied for decades, the molecular pathways, etiology and pathogenesis of these diseases remain poorly understood.

Although the placentas of no two mammalian species are the same (Carter, 2007), the placentas of humans and mice have strong similarities (Adamson *et al*, 2002; Georgiades *et al*, 2002). In both species, maternal blood from the uterine arteries enters the placenta from large diameter, spiral arteries located in the maternal decidua. The maternal blood then percolates through a dense mesh of channels created and lined by fetal trophoblast cells in which an equally dense network of fetal capillaries is localized. This region is the site of feto-maternal exchange and is called the *villous tree* in humans and the *labyrinth* in mice. In both species, the umbilical vessels connect the fetal capillaries of the placental exchange region with the fetal body circulation (Georgiades *et al*, 2002). Earlier in gestation in the mouse, the yolk sac plays a major role in feto-maternal exchange, but this role is taken over by the placenta later in pregnancy. In humans, the yolk sac is largely vestigial and the chorionic villi play an exchange role from early stages.

A critical region of the placenta and the focus of our study is the feto-maternal exchange tissue of the placenta; the human villous tree and the mouse labyrinth (Rossant and Cross, 2001). In both species, this region is involved in the transport of gases, nutrients and waste products between the maternal and fetal circulations throughout the majority of pregnancy and is therefore critically important to sustain normal fetal growth and development. From a pathological perspective, this region is of high interest because it is abnormal in severe early onset PE with IUGR in human pregnancies (Egbor *et al*, 2006). The placental exchange region of these patients has a reduction in the volume and surface area of the villous tree. Clinically, this group has the highest risk of serious maternal and fetal morbidity and mortality. Finally, given the extremely large surface area for contact with maternal blood, it is likely that tissues in this region will be an important site for production and release of bioactive factors and potential biomarkers into the maternal circulation.

A recent microarray analysis of the development of the mouse placenta has revealed some of the processes guiding placenta evolution (Knox and Baker, 2008). Several large-scale microarray datasets for various mouse and human tissues including the placenta have been published (Su *et al*, 2004; Zhang *et al*, 2004), yet there has not been a detailed analysis of co-expressed orthologs in human and mouse tissues on a proteome level. We have shown earlier the utility of shot-gun proteomics to survey protein expression in diverse mouse tissues; this method can be used to define tissue specificity and predict protein subcellular localization (Kislinger *et al*, 2003, 2006; Cox *et al*, 2007; Gramolini *et al*, 2008). In this paper, we present a comparative proteomic and transcriptomic investigation of microdissected placental tissue from the mouse labyrinth and the human villous tree at or close to term. We show over 70% co-expression of orthologous genes in both species. More interestingly, over 80% of genes known to cause placental phenotypes in mouse are co-expressed in both species. The human dataset is further validated by a large-scale comparison of immunohistochemistry (IH) data of over 4000 gene expression patterns in the human placenta available from the Human Protein Atlas (Ponten *et al*, 2008). A subset of the genes form a tight protein–protein interaction network involving 15 of these phenotype-associated genes and 34 novel candidate proteins also likely important in placental structure and/or function.

Results

Two term (38 weeks) human placentas were microdissected to harvest eight villous trees. These formed two biological replicates that were each split into three technical replicates for protein analysis and one replicate for microarray analysis. Six litters of C57Bl/6J inbred mice were dissected at E17.5 (one day before normal delivery to ensure clean sample collection) on two occasions, placentas were removed and the labyrinths were microdissected. These formed two biological replicates that were then split into three technical replicates each for protein analysis and one replicate each for microarray analysis (i.e. two groups of ∼25 mouse labyrinths were pooled for each analysis) (see Materials and methods).

Six subcellular fractions were generated from the tissue samples, they were cytosol, salt extracted nuclei, Triton-X-100 extracted nuclei, soluble mitochondrial proteins, Triton-X-100 extracted mitochondrial proteins and microsomes (typically a collection of plasma membrane, golgi, endoplasmic reticulum and other intracellular vesicles).

Protein fractionation efficiency for both mouse and human samples were validated by western blotting and showed good separation of cellular compartments (Supplementary Figure 1). Subcellular fractionation has been used by us and others as a means of increasing the coverage and biological information of the proteome versus whole cell extracts (Kislinger *et al*, 2003; Schirmer *et al*, 2003; Hoffmann *et al*, 2005; Cox and Emili, 2006; Foster *et al*, 2006). In total, five replicates for each of the six cell fractions from human and mouse were subject to MudPIT analysis and database searching, statistically filtered by peptide and protein, quantified by spectral counting and then normalized (see Supplementary information).

Microarray data of replicates for both species showed good correlation of signal intensities between replicates (Pearson correlation score of 0.998 for human and 0.992 for mouse) with an average of ∼18 000 probe sets reported as present for both species. Protein analysis identified 4612 and 4662

proteins in the human and mouse feto-maternal exchange tissue, respectively (Supplementary Tables I and II) representing 4226 and 4111 genes, respectively. The estimated false discovery rates based on matches in the decoy database were ~1–2% on the protein level (62/4612 and 40/4662 for human and mouse, respectively; Supplementary Table III). MudPIT analysis requires replicate analysis to achieve saturation of detection. This is not to say that all protein isoforms have been detected but that the limits of separation chemistry and mass analysis have been reached. A clear saturation of detection was observed as >95% of all proteins were detected in two or more replicates in each cell fraction from each species (Supplementary Figure 2). This indicates that further replicate analysis, either biological or technical, would not substantially increase the numbers of identified proteins.

Fractionation efficiency was further validated by hierarchical clustering of the data, which resolved the observed proteins into groups enriched to distinct subcellular regions. Each of these groups were tested for statistical enrichment of Gene Ontology (GO) cellular locations and molecular function terms using the GOFFA feature (Sun *et al*, 2006) of ArrayTrack and all were found to have a significant enrichment ($P < 0.01$ and fold enrichment >2) for terms expected for the enriched subcellular region (Supplementary Figure 1).

As neither manual inspection of a statistically significant proportion of the ~10 million mass spectrometry (MS)/MS scans nor large-scale western blotting are feasible methods of data validation, we sought an alternative method of data validation. The Human Protein Atlas (www.proteinatlas.org) is establishing an image database of different human tissues stained with antibodies against each protein in the human genome (Ponten *et al*, 2008). Currently, Protein Atlas contains 5062 genes in which at least one antibody had been generated (January 2009) and for which staining had been performed on the trophoblast portion of the placenta. Of these, 4146 genes have been scored as positive staining in the trophoblast (≥weak staining). We extracted this data along with Ensembl gene identifiers and antibody quality control data. All antibodies at Protein Atlas have quality control scores associated with them that include an overall interpretation of the IH staining as being supported by the literature. Additionally, western blots (WBs) are used to determine specificity of the staining (single or multiple bands) and the molecular mass of the target protein as compared with literature. IH is ranked as high, medium, low and very low confidence, whereas WBs are ranked on a scale of 1 (highest confidence) to 7 (lowest confidence).

We generated a table of Ensembl gene IDs with representative IPI terms and all associated human placental proteomics and microarray data from this study (18 760 genes). We mapped this current dataset to the antibody dataset, which created a final table of 4872 genes for which proteins were assayed by ourselves and/or Protein Atlas. We set up a contingency table to calculate: true positives (MS and IH agree that a protein is present), false positives (MS detected the protein but IH does not), false negatives (MS does not detect the proteins but IH does) and true negatives (MS and IH do not detect the protein). From these values, we calculated the precision (true positives over the sum of true and false

positives) and recall (true positives over the sum of true positives and false negatives) of our data versus the Protein Atlas results (Table I). Precision and recall are becoming the preferred metrics for judging a classification problem (Lu *et al*, 2004); expressed versus not expressed. We also generated contingency tables for microarray data versus IH results (Table I). There are two potential caveats of this comparison. First, IH methods are imperfect at detecting every protein present in the feto-maternal exchange tissue. Second, we assessed all cell types in a microdissected region of the human placenta, whereas the Protein Atlas assessed trophoblast from a whole placental specimen.

On the basis of this analysis, the precision of our data is very good with a value of 0.88 (1444/1649) for all 4872 genes, whereas the recall was relatively poor with a value of 0.36 (1444/3989). When we included only high quality antibody data (i.e. 1426 genes detected using antibodies with a medium or high IH ranking and a WB score of 1, 2 or 3), this substantially improved the recall to 0.58 (699/1199) and the precision improved slightly to 0.90 (699/777). The precision of the microarray was very similar to the MS data, but had better recall values using either all (0.66) or high quality antibody data (0.78). The high correlation (high precision) between the datasets indicates that both methods have low false positive detection. The lower recall is likely the result of the random sampling and sensitivity of the technique, a common problem with global shot-gun proteomics, which still fails to identify entire proteomes (Washburn

Table I Precision and recall statistics to compare expression of 4872 proteins in our dataset (MS or microarray data) to expression detected by immunohistochemistry in Protein Atlas

All antibodies		Immunohistochemistry		Precision
		Exp	NE	
	Total	3989	883	
MS				
Exp	1649	1444	205	0.88
NE	3223	2545	678	
Recall		0.36		
mRNA				
Exp	3033	2640	393	0.87
NE	1839	1349	490	
Recall		0.66		
High quality antibodies		Immunohistochemistry		Precision
		Exp	NE	
	Total	1199	227	
MS				
Exp	777	699	78	0.9
NE	649	500	149	
Recall		0.58		
mRNA				
Exp	1031	934	97	0.91
NE	395	265	130	
Recall		0.78		

Exp, expressed; NE, no expression; MS, mass spectrometry protein data; mRNA, microarray transcript data.
Also shown is the same comparison using only the high quality immunochemistry data from Protein Atlas (available for 1,426 proteins).

et al, 2001; Wolters *et al*, 2001; Liu *et al*, 2004; Zybailov *et al*, 2005).

Human-mouse orthologs

Developing an appropriate model of a human disease requires identifying functional conservation of candidate human disease genes in a model organism. Typically, this has been done on a case-by-case basis or implicitly assumed due to the high conservation of the human and mouse genomes.

To systematically assess the molecular similarity of human and mouse feto-maternal placental exchange tissue, we built a database of protein and microarray data linked to Ensembl gene IDs for both species. A human to mouse ortholog table was obtained from Ensembl through Biomart (www.ensembl. org/biomart/martview) and used to link our datasets. Orthologs are classified as one-to-one, one-to-many or many-to-many by generating single linkage clusters of gene families of different species based on best reciprocal BLAST scores. In this way if a protein from each species is found at the end of a node they are one-to-one orthologs. If a single protein from one species is found in a node with two or more proteins from another species these would be one-to-many orthologs, as the paralogs in the second species have sufficient similarity to the single protein in the first species that they cannot be separated further. Lastly, if there are multiple proteins from both species at the end of a node then they constitute a many-to-many ortholog group. Genetic synteny is not considered in the generation of these ortholog tables.

As a first assessment of the molecular similarities between both species, we determined the co-expression of mRNAs of one-to-one orthologs in human villous and mouse labyrinth. One-to-one orthologs were chosen as they represent the simplest scenario for gene ortholog comparisons and a valuable first step for comparative analyses. Although these orthologs contain housekeeping and metabolic genes, many signaling proteins and transcription factors are represented. There were 12 941 one-to-one ortholog genes with probe pairs in human and mouse represented in our data. We tested the correlation of the two datasets based on the detection of the transcript in both tissues. We summed the detection score of the two replicate arrays, where P=1, M=0.5 and A=0 (i.e. present, marginal, absent calls generated by Affymetrix GCOS software), for both species. This analysis is graphically displayed in Figure 1A, showing a high degree of concordance between detected transcripts for both species. Briefly, 6609 transcripts had present calls in both replicates in both species. A total of 2869 transcripts were determined to be uniquely expressed between both species, 1804 in mouse and 1065 in human, and 2567 transcripts were recorded as absent in both human and mouse.

As a measure of correlation, a Jaccard score (see Materials and methods) was used, which gave an average score of 0.71 (where 1.0 would be a perfect correlation and 0 no correlation) for the comparison of the two mouse replicates against the two human replicates. This seemed to indicate a high degree of molecular similarity; however, no published large-scale comparisons of human and mouse orthologs are available in the literature as a measure of comparison. We therefore assembled a large set of microarray data for 54 different mouse tissues from the GEO repository plus mouse labyrinth from this study (Supplementary Table IV), and four different human tissues (Ge *et al*, 2005) (brain, heart, spleen and kidney) from the GEO repository (GSM44671, GSM44673, GSM44675, GSM44690) and human villous tree from this study. The correspondence between the assembled mouse and human datasets was assessed using the Jaccard function as a measure of correlation. The distribution of Jaccard scores for various mouse tissues obtained for each of the five human tissues is graphically displayed in Figure 1B. In all cases, the corresponding tissue from mouse was returned as the top matching score (highlighted in red in Figure 1B). In the case of the villous tree, the mouse labyrinth was first with the whole placenta a close second, which would indicate that microdissection has enriched the sample. The distribution of Jaccard scores for the villous sample seems to reflect a generally greater level of molecular similarity with the other mouse tissues as compared with the other four human tissues. This bias could be due to the use of a different microarray set for the villous sample (this study) compared with the other tissues (from GEO).

Combined analysis of transcription and translation

We next combined the microarray and proteomic data to generate a direct survey of both tissues on the transcriptional and translational level. This integrative analysis is beneficial as it enables the estimation of missing data and allows the determination of experimental artifacts in organ analysis such as blood proteins (Cox *et al*, 2005, 2007; Kislinger *et al*, 2006). To reduce redundancy, we collapsed all protein and probe IDs to Ensembl gene IDs allowing only a single representative IPI or Affymetrix probe ID to be used (see Materials and methods). This lowered the total number of proteins from 4662 to 4226 and 4612 to 4111 for mouse and human, respectively. After mapping to the microarray data, a total of 4161 and 4061 protein and array probes were linked for mouse and human, respectively. Only a small fraction of gene products had an absent call in both RNA microarray replicates and detectible levels of protein; 3% (134 genes) for mouse and 6% (277 genes) for human. As anticipated, a large number of these were blood proteins and extracellular proteins that may have been generated elsewhere in the body or perdure long after mRNA has been turned over. Conversely, there were a large number of protein/probe pairs for which no protein was observed despite present calls in both RNA microarray replicates. Protein/probe pairs for orthologous proteins that were observed in both species (cluster I), one species (cluster II) or in neither species (cluster III) were independently clustered (Figure 2A). Organellar localization associated with significant enrichment for appropriate GO annotation terms was observed for cluster I proteins (Figure 2B). Interestingly, mouse versus human organellar localization showed striking agreement for co-expressed ortholog proteins in cluster I (Figure 2A and C) as predicted by Pearson correlation, which argues for conservation of function for most ortholog proteins in this cluster.

In cluster II, proteins are missing from one species but message was detected in both (Figure 2A). This was likely caused by the failure to detect proteins due to *random sampling*, a problem also observed in prior large-scale
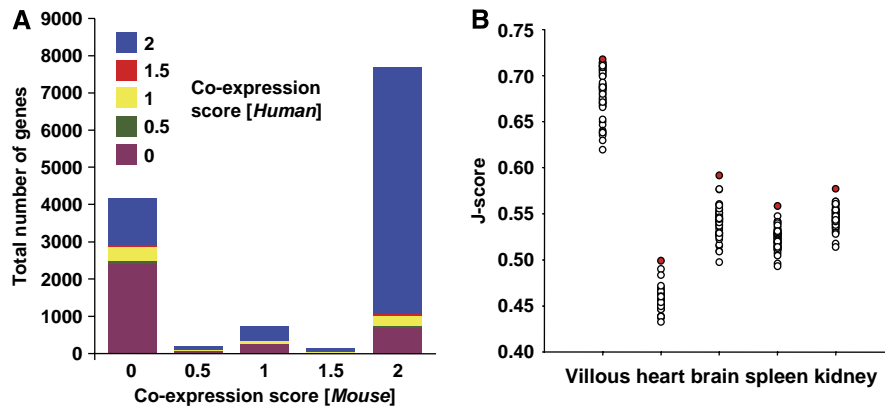
Figure 1 Correlation of microarray expression in human and mouse. (**A**) A stacked bar graph demonstration of the high correlation between human and mouse microarray expression. The duplicate arrays were scored as present (1), marginal (0.5) or absent (0) and summed. Excellent agreement is observed for genes with two present calls in mouse and human with over 70% of all detected probes correlating expression in human and mouse. Furthermore, the majority of probes with absent calls also agree in human and mouse. (**B**) A scatter plot distribution of Jaccard scores for 5 human tissues against 55 mouse tissues is represented. Data were obtained from the GEO repository and J-scores were pair wise calculated. In each case, the corresponding mouse and human tissues always showed the highest correlation (solid red circle) as compared with the other 54 tissues (open circles).
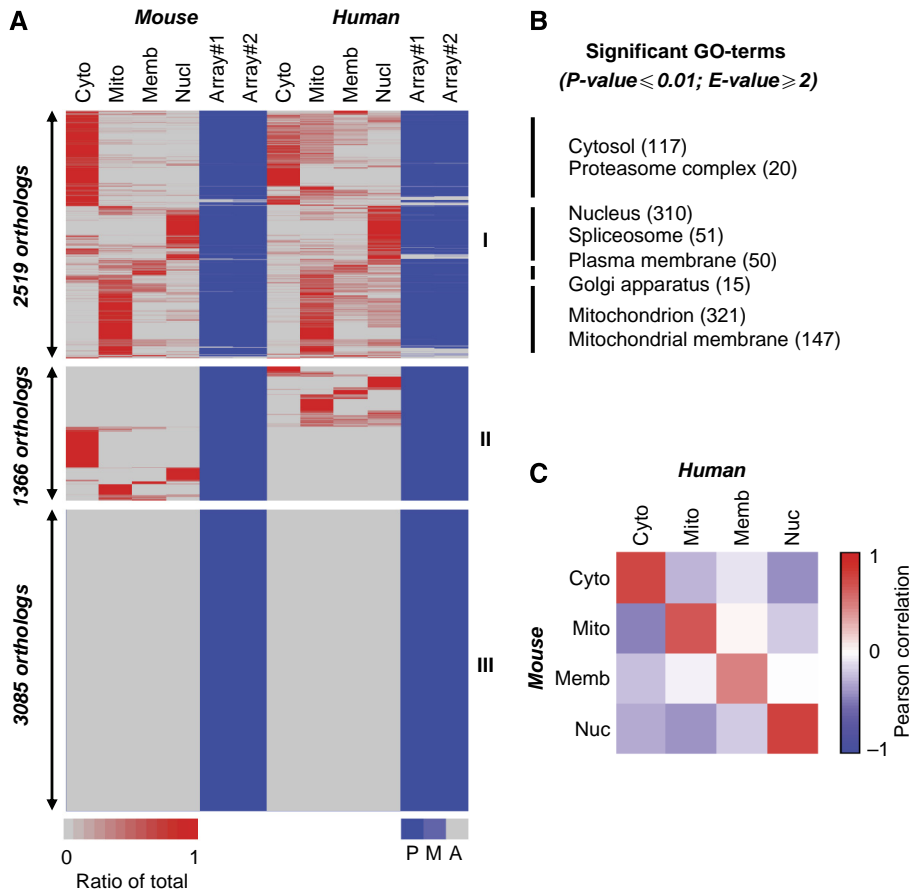

Figure 2 Heat map of protein and mRNA ortholog expression in human and mouse. (**A**) A heat map of one-to-one orthologs showing protein expression (in red with scale shown at left below) and mRNA expression (in blue with scale shown at right below) for mouse and human placental exchange tissues. All 6970 confidently detected ortholog proteins are shown clustered into three groups; proteins observed in both species (cluster I), one species (cluster II) or in neither species (cluster III). (**B**) Significant enrichment for appropriate Gene Ontology annotation terms was observed for cluster I proteins. (**C**) The Pearson correlation (mouse versus human) for organellar localization of co-expressed ortholog proteins in cluster I. High agreement suggests conservation of function for most ortholog proteins in this cluster.

proteomics studies (Liu *et al*, 2004). Limitations in detector sensitivity and/or separation chemistry of peptides have been proposed as reasons for the inability of proteomics studies to

detect the entire proteome. Protein instability, short half-life and/or low translation efficiency could be other potential explanations. In the case of cluster III, in which protein was not

detected in either species, but mRNAs were observed in both replicate arrays for human and mouse, several explanations are possible. First, it is possible that these genes are poorly translated from their transcripts as part of a gene regulatory mechanism or they are of very low abundance, impeding detection by MS.

To further investigate this observation, we obtained a representative IPI accession matching the corresponding Ensembl gene for cluster III and then performed a comparison *in silico* of the physicochemical properties of peptides observed in the co-detected group (cluster I) to those that should have been detected in cluster III. We calculated and compared the peptide hydrophobicity, isoelectric point and tryptic peptide mass distributions and were unable to detect significant differences between both clusters (Supplementary Figure 3), suggesting that lack of detection of the proteins was not a technical artifact.

We then compared the microarray probe signal intensity for genes detected in all three clusters (Figure 3A). Microarray probe intensities in cluster I were significantly stronger than

clusters II and III ($P<0.001$, two tailed T-test assuming unequal variance), implying a generally stronger expression for this set of orthologous genes. Lower mRNA expression in clusters II and III may result in low levels of protein translation, which could explain why proteins in these clusters were not consistently detected by MS-based proteomics.

Next, we compared the proportion of proteins with GO-term annotations in clusters I and III. The number of proteins with no GO-term annotation was increased in cluster III (Figure 3B), indicating that proteins in this cluster are less well characterized possibly because they are harder to detect or because they are not transcribed. To evaluate the detectability of proteins in both clusters and to obtain additional evidence for the translation of genes in cluster III, we compared our data with several large-scale mouse proteomic studies (Foster *et al*, 2006; Kislinger *et al*, 2006; Adachi *et al*, 2007; Gramolini *et al*, 2008; Graumann *et al*, 2008) and one unpublished dataset (TK personal data). We observed that $\sim 52\%$ of proteins in cluster III were not detected in prior studies compared with only 3.5% in cluster I (Figure 3C).

To obtain an independent assessment of this data, we again used the human placenta IH data at Protein Atlas (Supplementary Table V) (Ponten *et al*, 2008). For cluster I proteins, we found a precision of 93% for all 834 genes with IH data (i.e. detected by both), whereas for cluster III proteins, precision was only 13% of 711 genes, which indicates that there are proteins present for cluster III but they go largely undetected by MS-based proteomics. However, there were twice as many genes with no observed expression by IH in human placenta in cluster III (111 or 13.5%) than in cluster I (58 genes or 6%), indicating that cluster III may be enriched with genes for which only transcripts are present.

In summary, our data suggests that at least a part of the gene products of cluster III are expressed at lower levels or only in a defined subset of the cell types present in the analyzed tissue, in either case they are at a concentration below the detection limit of data dependent shot-gun proteomics. There still remains the possibility that a portion of cluster III proteins are differentially regulated at the translational rather than transcriptional level. Further experiments are needed in the future to specifically investigate this observation.

A major goal of this study was to identify a core panel of co-expressed ortholog proteins in human and mouse feto-maternal placental exchange tissue that could be used to study placenta biology in functional mouse models or used to develop biomarkers for placental diseases. Thus, we will focus our attention on the 2519 proteins co-detected in both species (cluster I).
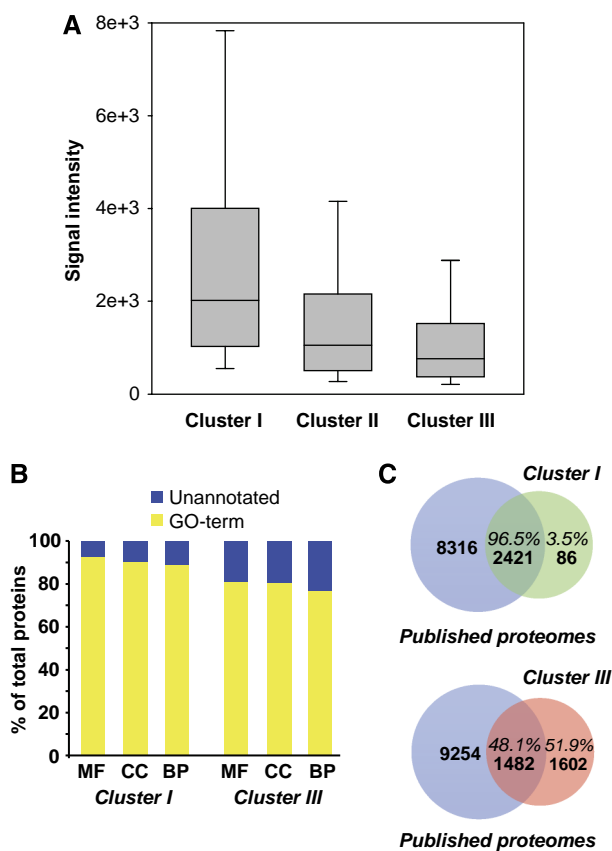

Figure 3 Comparison of ortholog protein clusters. (**A**) Shown is a box chart of mouse microarray signal intensities for genes in the three distinct ortholog expression clusters. Outlines of the box are the 25th and 75th percentile with the median shown as a line. Error bars represent the 10th and 90th percentile. (**B**) ProteinCenter software suite was used to annotate cluster I and III with Gene Ontology terms (MF, molecular function; CC, cellular component; BP, biological process). Cluster III has a significantly smaller percentage of proteins with GO annotation than cluster I (**C**) Protein accessions from cluster I and III were mapped against six other large-scale proteomics datasets containing over 10 000 unique proteins (published proteomes, blue) using ProteinCenter. The number and percentage of proteins that do or do not overlap with the published proteome are shown for cluster I (green) and cluster III (orange).

Species unique expression

In all, 2869 one-to-one orthologs were uniquely expressed in each species, which highlights the level of divergence in gene expression between human and mouse since the evolution of the placenta. We analyzed the gene expression data to gain insight into this evolutionary process (Supplementary Table VI). There are a number of factors that drive placental evolution that are related to the number of fetuses that need to be supported, the relative sizes of the mother and fetus and gestational time (Knox and Baker, 2008). We constrained the

comparison to include only uniquely expressed one-to-one orthologs in which protein and mRNA expression correlated (i.e. ortholog protein and mRNA detected in mouse, but the corresponding human protein was not detected and the Affymetrix microarray gave an absent call). This step is important, because missing data in proteomics could be the result of either true absence of the proteins or simply failure of detection (Liu *et al*, 2004). By combining both technologies, we are able to minimize this caveat. As examples, we note that *Notch2* and *Cyp19A1* were both detected only in the human, and *Cited1* was detected only in the mouse (Supplementary Table VI). These are interesting candidates as both the Notch family of proteins and the transcription factor *Cited1* have been shown by mutational analysis to be important for mouse placental development. Estrogen in humans is critical for pregnancy; therefore, the expression of *Cyp19A1*, the major estrogen aromatase in each species is of importance.

Notch signaling is important for many aspects of placental development as shown by mouse knockout models (Gasperowicz and Otto, 2008). Specifically, *Notch2* plays a role in the formation of the maternal blood sinus of the placenta in mice (Hamada *et al*, 2007). The expression of *Notch2* seems to both become decreased and greatly restricted by E11.5. Our detection of *Notch2* uniquely in the human villous sample suggests a change in the role of this receptor in placental development as its expression has persisted in term human placentas but appears to be absent in mouse by this stage. Notch 1 and 4 are implicated in fetal angiogenesis in the placenta (Gasperowicz and Otto, 2008), and we noted expression of *Notch1* and *Notch4* in both human and mouse samples, indicating conservation of their role during placentation.

Cyp19A1 is the major aromatase for estrogen synthesis and is believed to be important for endometrium and placental development in humans (Simpson *et al*, 1997). Mouse *Cyp19A1* homozygous null embryos show no changes in placental development but maternal effects are not known because homozygous females are infertile (Fisher *et al*, 1998). In mouse models of endometrium formation, estrogen is dispensable for decidualization (Kaitu'u-Lino *et al*, 2007). The absence of *Cyp19A1* expression in the mouse labyrinth supports a divergent role for estrogens in the regulation of implantation and placental development in the human as compared with mouse.

Cited1 is essential for normal formation of the labyrinth layer of the mouse placenta (Rodriguez *et al*, 2004). We noted that only mouse showed gene expression of *Cited1*, whereas both human and mouse placentas show gene expression of *Cited2* and *Cited4* by microarray. Our data indicate that the role of *Cited1* may have diverged, but may be compensated for by the conserved expression of *Cited2* and *Cited4* expression in human and mouse. Alternatively, the role of *Cited1* may be mouse specific, for example, used by sinusoidal giant cells in the generation of maternal blood spaces in the labyrinth, whereas there are no giant cells lining maternal blood spaces in the human placenta.

Complex orthologs

Genes with complex one-to-many and many-to-many mappings represent genes that have undergone duplications after

speciation of mouse and human. This complicates mapping of homologs and orthologs. However, there are several interesting examples in this group of proteins. One example is the cathepsin peptidases (Table II). In mice the cathepsins are expressed in the trophoblast and developing placenta, but seem to have a high degree of functional redundancy complicating their molecular dissection in placental development (Deussing *et al*, 2002). However, they most likely function to remodel the extracellular matrix during trophoblast invasion and possibly during vascularization. Cathepsins are also expressed in the human placenta. Although it is difficult to determine how they relate to the expression and function of the mouse cathepsins (Mason, 2008). The cathepsins have undergone multiple duplications before and after speciation of mouse and human (Deussing *et al*, 2002). The cathepsin L (CTSL) gene in humans is expressed in the placenta and is detected in our study by both protein and microarray. This gene seems to have undergone several rounds of duplication in the mouse. Table II shows that seven of the eight mouse CTSL orthologs have detectable expression by microarray and five of eight also have detectable protein in our study. This would argue that the orthologs of human CTSL in mouse have been selected to keep placental *cis* regulatory elements intact (Table II).

Conservation of phenotypic proteins

The dataset generated in this study provides a resource for the informed selection of candidate genes to be evaluated in mice as models or as biomarkers of human placental disease. In this context, we mined the OMIM repository (http://www.ncbi.nlm.nih.gov) for human placental disease genes using the specific search terms *placental insufficiency* and *intrauterine growth restriction*. On the basis of this analysis, we located seven genes, known or suspected to cause placental abnormalities in humans. Interestingly, all seven genes were co-expressed in the mouse labyrinth, and prior work showed placental phenotypes when these genes were mutated (Table III). Of these seven genes, five were in cluster II, in which proteins were found only in one species but transcripts were detected in both and two were found in cluster I.

Encouraged by this result, we systematically mined our data of co-expressed one-to-one orthologs for genes with known placental phenotypes in mouse to guide the discovery of novel placental disease targets in humans. The number of validated phenotypic mouse genes vastly exceeds those in human due to systematic deletion efforts in mouse (Skarnes *et al*, 2004; Forrai and Robb, 2005; Nord *et al*, 2006). To accomplish this goal, we made use of the large number of well-annotated mouse phenotypes available at MGI (http://www.informatics.jax.org). Specifically, 183 one-to-one orthologous genes with mutant alleles are annotated in mouse to cause *abnormal placental morphology* (MGI Mutant Phenotype Database, January 2008). We detect 170 of these in our study; the majority of these (138 unique genes) are co-expressed. Of the co-expressed genes, more than half (81 genes) were detected by MS-based proteomics in at least one species (Supplementary Table VII; Figure 4).

Several other more specific phenotype ontologies were observed in our dataset, most notably phenotypes of the

Table II Comparison of expression of cathepsins genes of one-to-many and one-to-one orthologs

	Mouse			Human			
Gene Symbol	Total Protein	Array 1	Array 2	Gene Symbol	Total Protein	Array 1	Array 2
One-to-many							
Cts3	34	P	P	CTSL1	73	P	P
Cts6	117	P	P				
Cts7	0	A	A				
Cts8	0	P	P				
Ctsj	160	P	P				
Ctsm	3	P	P				
Ctsq	395	P	P				
Ctsr	0	P	P				
One-to-one							
Ctsa	105	P	P	CTSA	37	P	P
Ctsb	869	P	P	CTSB	237	P	P
Ctsc	11	P	P	CTSC	142	P	P
Ctsd	291	P	P	CTSD	388	P	P
Ctse	0	A	A	CTSE	0	M	P
Ctsf	7	P	P	CTSF	0	P	P
Ctsg	0	A	A	CTSG	26	A	A
Ctsh	3	P	P	CTSH	3	P	P
Ctsk	0	P	P	CTSK	0	P	P
Ctsl	12	P	P	CTSL2	0	P	P
Ctso	0	P	P	CTSO	0	P	P
Ctss	0	P	P	CTSS	2	P	P
Ctsw	0	A	A	CTSW	0	A	A
Ctsz	56	P	P	CTSZ	11	P	P

Shown are the CTSL ortholog family and the other one-to-one cathepsin orthologs with total spectral counts and absent (A) or present (P) calls from the replicate microarrays.

Table III List of all genes known or suspected to cause placental abnormalities in humans

Human symbol	Cluster	Reference	Human placental phenotype	Mouse symbol	Reference
CDKN1C	II	Hatada *et al* (1996)	Enlarged placenta	Cdkn1c	Takahashi *et al* (2000)
CUL7	II	Maksimova *et al* (2007)	Intrauterine growth retardation and vascular abnormalities	Cul7	Arai *et al* (2003)
IGF2	I	Qiu *et al* (2005)	Placenta over growth	Igf2	Constancia *et al* (2002)
GPC3	II	Veugelers *et al* (2000)	Placenta over growth	Gpc3	Chiao *et al* (2002)
THBD	I	Franchi *et al* (2001)	Possible association with placental insufficiency in humans	Thbd	Healy *et al* (1995)
AKT1	II	Qiu *et al* (2005)	Possible association with IUGR	Akt1	Yang *et al* (2003)
GRB10	II	Yoshihashi *et al* (2000)	Possible association with IUGR	Grb10	Charalambous *et al* (2003)

All were detected in both the mouse labyrinth and the human villous tree. All have known placental phenotypes in mice. Cluster denotes the cluster number from Figure 2A where the expressed gene products are observed.

labyrinth and the placental vasculature (Supplementary Table VII; Figure 4). This panel of placental phenotypic orthologs co-detected in both organisms in our study represents a class of readily identifiable markers of placental development and disease. The observation of mouse-human co-expression of a substantial subset of genes that cause placental phenotypes when mutated in the mouse, strongly implicates these genes as candidates for causative factors in human placental diseases.

Detection of phenotypic protein networks

The placenta is a comparatively under-studied tissue, for example, a PubMed search of 'heart and human' yields > 500 000 original research articles, whereas 'placenta and human' finds < 50 000. For this reason, it is to be expected that a large number of genes that cause placental abnormalities likely remain undiscovered. To expand on our panel of candidates, we integrated our protein dataset with protein–protein interaction data as a strategy to identify additional proteins likely to be important in placental physiology and disease (Kim *et al*, 2008; Pena-Castillo *et al*, 2008). The core network was limited to the set of 2519 co-expressed orthologous proteins identified in cluster I. Their interactions were assembled by a query against the I²D protein–protein interaction database ver 1.71 (Interologous Interaction Database; http://ophid.utoronto.ca/i2d; Brown and Jurisica, 2007). Interacting proteins were next mapped against the phenotype ontologies, *abnormal placenta labyrinth morpho-*
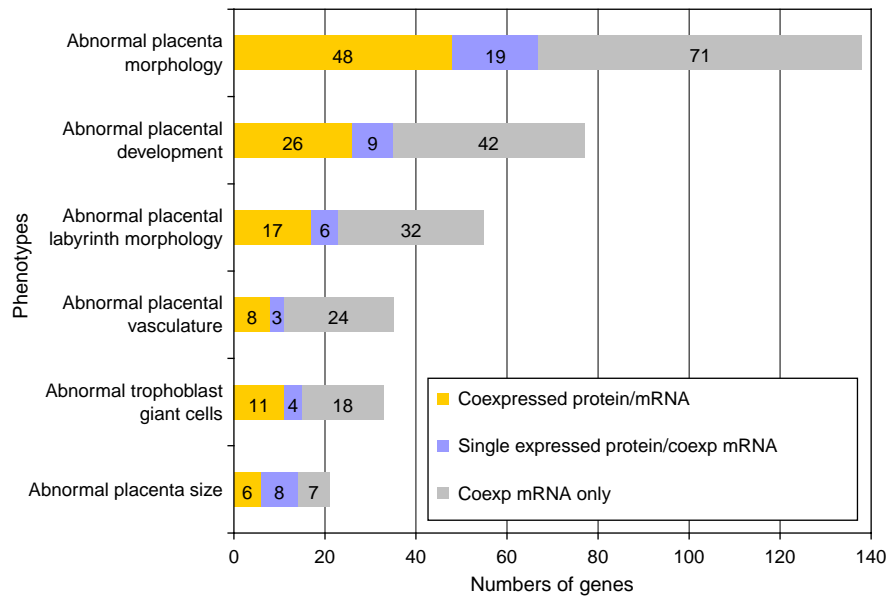
Figure 4 Orthologs with annotated placental phenotype. The numbers of phenotypic gene orthologs identified by microarray and protein. Placental phenotype ontology terms are from MGI (http://www.informatics.jax.org). Numbers in each phenotype category are broken down into cluster I (co-detected by protein and microarray), cluster II (unique detection by protein and co-detection by microarray) and cluster III (co-detection by microarray only).

logy (MP:0001716) and *abnormal placenta vasculature* (MP:0003231) available at MGI. Those that mapped were used as seed proteins. We then extracted a subnetwork of interacting proteins that showed significant association with these seed proteins (P-value $\leqslant 0.05$) but were not themselves annotated to the above phenotype ontologies. The rational of this integration was to obtain novel proteins, likely to be important to placental biology, but currently lacking direct phenotypic association to the highly vascular labyrinthine feto-maternal exchange region. The final network consisted of 15 seed proteins and 34 additional interacting proteins that had statistically significant associations (see Materials and methods, Supplementary Table VIII).

Although we enriched our samples by microdissection, they are still composed of multiple cell types. This may confound our network analysis as not all members of the interaction network may be co-expressed in all cell types present. We again used the Protein Atlas IH resource to classify the proteins as to what cell type(s) they are expressed. We limited our cell types to the syncytial trophoblast, cytotrophoblasts, mesenchyme and endothelial cells. The majority of the network could be mapped into Protein Atlas (41/49 proteins had antibodies in Protein Atlas). Of these, 36 were scored at Protein Atlas as positive for expression and five were scored as negative (Cd82, Itga2b, Icam1, Itgb3, Dcn). We manually inspected the images of all 41 proteins and found that of the five proteins scored as negative Itga2b and Cd82 appeared negative, Icam1 was present in the endothelial cell only, whereas Dcn and Itgb3 were expressed in the mesenchyme of the villous tree (Supplementary Figure 4). Although the image from Protein Atlas shows expression of Cd82 in the decidua only, an earlier publication found expression in the villous mesenchyme (Gellersen *et al*, 2007). The expression patterns are summarized in a Venn diagram and displayed along with representative images from Protein Atlas (Figure 5).

Of the 39 proteins for which we could score expression in the IH images from Protein Atlas, we found 17 to be expressed in common, 10 were expressed in the syncytial and cytotrophoblast cells, 6 were expressed in the syncytial trophoblast and the endothelial cells, Crk and Fhl2 were unique to cytotrophoblasts and Icam1 was unique to endothelial cells. A network was drawn using NAViGaTOR (ver 2.15; http://ophid.utoronto.ca/navigator) and color coded to indicate in which cell type(s) the proteins were expressed. These data indicate a general protein–protein network that has cell-type unique aspects.

This network is enriched in cell adhesion, ECM interacting proteins and adherence junctions. Our analysis indicate that these networks are more enriched to the endothelial and syncytial trophoblast cells, whereas many of the cell-surface receptors and cell signaling adaptors and kinases are more enriched to the syncytial and cytotrophoblast cells. EGF signaling seems to play a prominent role in the network (Figure 6). This is supported by experimental evidence that shows that the labyrinth of *Egfr* knock-out mice are reduced in size and disorganized in an inbred 129/Sv background (Threadgill *et al*, 1995). *Smad4, Numb* and β-catenin are also included in the network, which could indicate a cross-talk of EGF signaling with Notch, Wnt and/or TGF-β/BMP signaling pathways.

As the goal of this work was to identify proteins with hitherto unknown involvement in the feto-maternal exchange region, we cross-referenced the network members back to mouse knockout phenotypes and OMIM to determine associations with human disease phenotypes. Of the 34 non-seed members of the network, 32 had reported mouse knockouts (*Cd82* and *Eps15* were the exceptions) and only one protein was associated with human disease (mutations in *Dcn* (Decorin) cause stromal corneal dystrophy) (Bredrup *et al*, 2005). Although these mouse knock-outs are not categorized
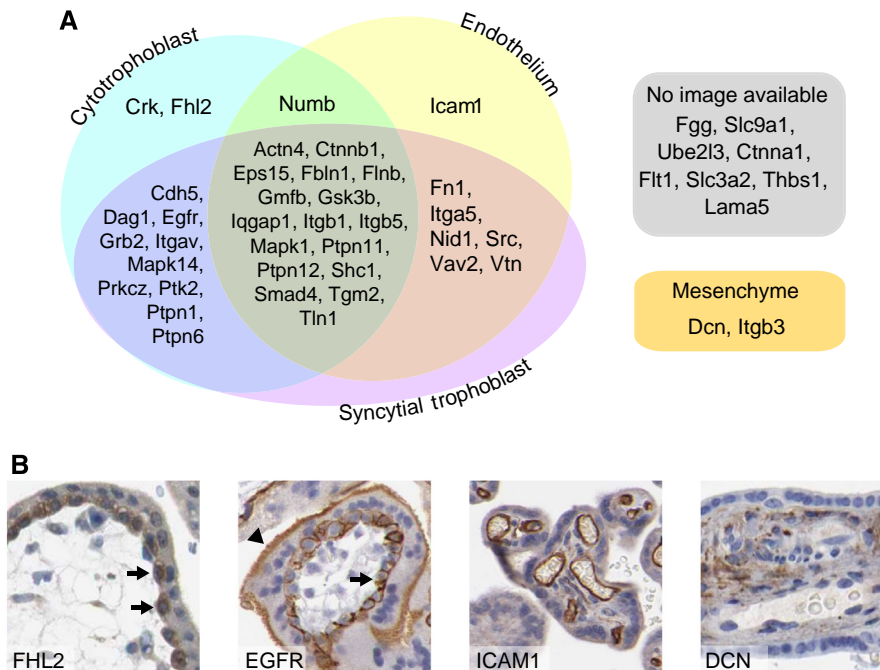
Figure 5 Cell-type specificity of the labyrinth phenotype protein–protein interaction network. Classification of cell-type expression of protein network members by immunohistochemistry data from Protein Atlas. (**A**) A Venn diagram of unique and overlapping expression patterns. (**B**) Representative images showing expression in the cytotrophoblast (arrows; FHL2), cytotrophoblast (arrows) and syncytial trophoblast (triangle; EGFR), endothelium (ICAM1) and mesenchyme (DCN).

as having placental phenotypes in the MGI database, careful examination showed that 12/34 non-seed network members are annotated has having embryonic lethality and three others with the annotation decreased litter size, hallmarks of as yet uncharacterized placental defects.

The largest protein family in our network is the integrins (*Itga2b, Itga5, Itgav, Itgb1, Itgb3* and *Itgb5*) (Figure 6). Integrins mediate cell adhesion to basement membranes and extracellular matrix, and cell signaling through fibronectin. Mutational analysis showed integrins alphaV, beta1 and beta3 are important in labyrinth and/or placental vascular development (Supplementary Table VII) so they were used as seed proteins in the network. Network analysis suggests integrins alpha5, alpha2b and beta5 are also likely involved (Figure 6). The phenotype of the *Itga5* knockout mouse further supports this suggestion because deletion causes vascularization defects in the embryo embryonic growth retardation and lethality between embryonic days 10 and 11 (Yang *et al*, 1993). Growth retardation is typically a sign of poor placental function and lethality at this time is typical of placental failure, as this is the stage at which the fetus becomes dependent on the placenta for exchange of nutrients and oxygen (Copp, 1995).

The functions of two proteins in the network have not been studied earlier using mouse mutagenesis; *Cd82* and *Eps15* (Figure 6). *Cd82* is a tumor suppressor and inhibits the spread of metastatic cancer cells (Dong *et al*, 1995). It forms interactions with *Darc* (Duffy antigen receptor for chemokines) to facilitate adherence of other cell types to endothelial cells (Bandyopadhyay *et al*, 2006). Our network shows interaction of *Cd82* with *Egfr* and *Itgb1*; both seed members of the network. *Darc* was only weakly detected in mouse labyrinth by only one of the two microarray replicates.

However, other proteins may facilitate a similar interaction of these cell types through *Cd82*. Cd82 is suspected to be involved in the repression of placenta invasion and appears to be upstream of Dcn expression in the decidua (Gellersen *et al*, 2007).

Eps15 is a target of activated Egfr kinase activity (Fazioli *et al*, 1993), and seems to be involved in negative feedback of *Egfr* signaling by promoting *Egfr* endocytosis and degradation (Fallon *et al*, 2006). Our network shows that *Eps15* interacts with *Egfr*, *Numb* and *Crk*. The available biological annotation of both *Cd82* and *Eps15*, and their interaction with other members of the network suggest a role in cell adhesion of trophoblast to endothelial cells and regulation of *Egfr* signaling, both of which are critical for placental development based on this study (Figure 6) and prior work (Threadgill *et al*, 1995). Thus, *Cd82* and *Eps15* are strong candidates for knockout models to determine their molecular roles in placental biology.

Discussion

An important question for the development of human disease models is the knowledge of functional conservation for candidate human disease genes in the model organism. In this study, we created a database of orthologous genes between human and mouse for the analogous tissue type, the feto-maternal exchange tissue of the placenta. These data have been assembled into an accessible and well-annotated database to provide a tool for clinician scientists interested in using the mouse to model human disease, and for biologists interested in building new hypotheses of development and evolution.
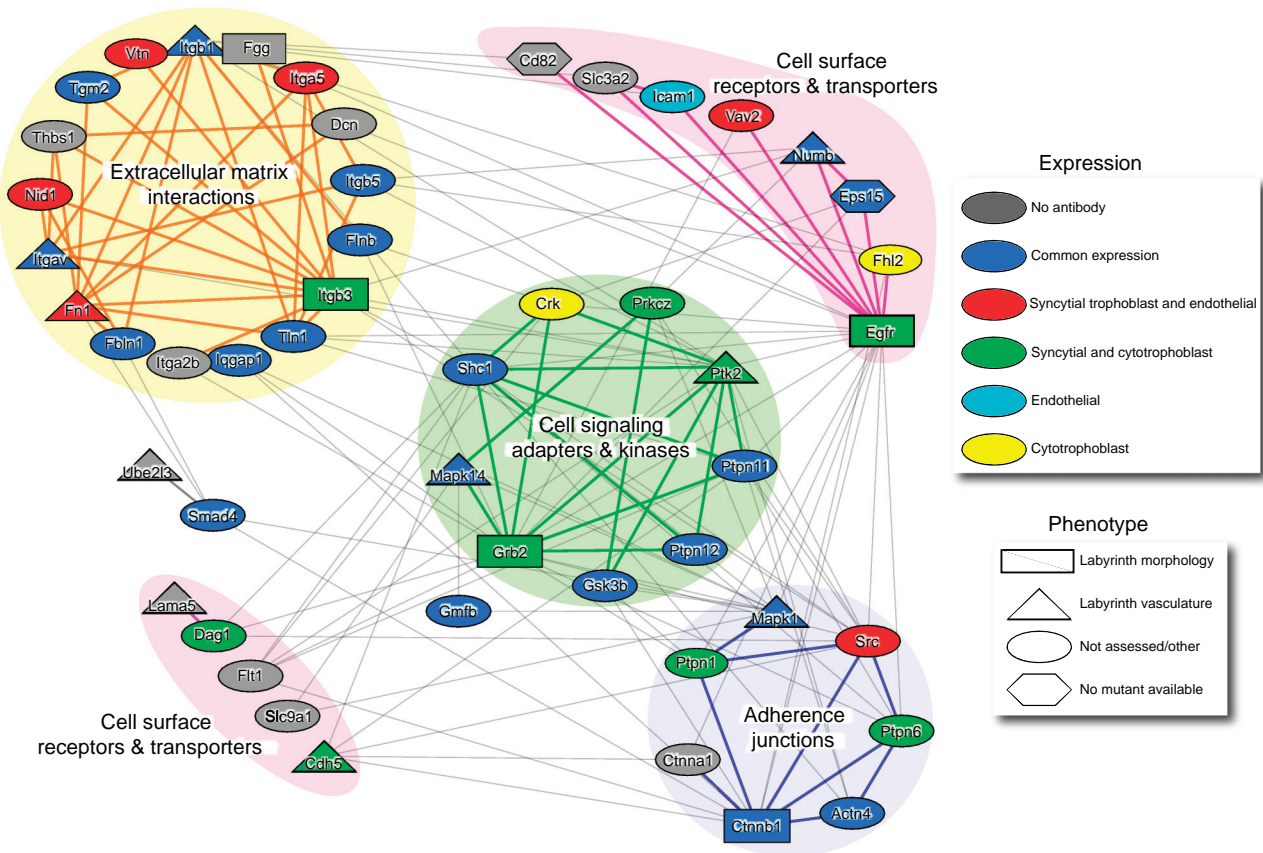
Figure 6 Schematic of the cellular topology of the protein–protein interaction network. Shown are protein–protein interaction networks from I²D version 1.71, seeded with proteins that give a placental labyrinth or placental vascular phenotype when knocked out (rectangle or triangle). Proteins are colored to show cell-type expression. Node shapes indicate the placental phenotype (rectangle or triangle) or if no placental phenotype is known (oval or hexagon). Enrichment of proteins involved in extracellular matrix (yellow circle) and focal adhesion complex (blue circle) were observed (P < 0.01). Other groupings are for clarity only. Eps15 and Cd82 (hexagons) are both novel members with no available knockout models. Visualization was done using NAViGaTOR version 2.15.

Despite our experimental and statistical rigor, our dataset does have some limitations. We have used human placenta samples from only two individuals (eight individual villi), which is not statistically representative of human placentas in general. Similarly, we used only one strain of mouse and placental phenotypes are dependent on genetic background. In addition, we only sampled one time point close to or at delivery in both cases. As it is extremely difficult to get preterm human placentas and delivered placentas typically have begun a necrotic process, the use of E17.5 mouse and cesarean delivered human placentas are reasonable compromises. Presumably, repetition of this study with additional human subjects and mouse strains may alter the dataset and reveal ethnic or genetic background dependent differences, which was not the goal of this study. Although we enriched our samples through microdissection, they still contain multiple cell types. We compensated for this by integrating the Protein Atlas IH of human placenta to obtain insight into the cell-type specific expression.

Prior work showed strong similarities between the proteome of the human and mouse RBC (Pasini *et al*, 2008). Similarly, a bioinformatic comparison between human and mouse survival genes (i.e. those that when deleted cause lethality before puberty or sterility) showed ∼80% conservation of phenotypic genes between species, suggesting that biological processes are well conserved (Liao and Zhang, 2008). This study reports the first direct experimental comparison of the proteome and transcriptome of a complex organ system in two mammalian species, and we report a similar degree of conservation (∼80%) of co-expressed phenotypic genes in human and mouse. In contrast to conservation, a recent comparative study of the post-synaptic density (PSD) of mouse and drosophila showed that core proteins are required for the generation of a PSD but that the addition of new proteins leads to changes and diversity among species (Emes *et al*, 2008). We found over 1000 one-to-one orthologs with expression in only one of the two tissues, which presumably may contribute to the anatomical and physiological differences between the same tissues of these two species.

Our insight into the similarities of human and mouse placental exchange tissue has created new opportunities for using the mouse model to study human development and disease. We identified a core set of conserved orthologs that were co-expressed, at the mRNA and/or protein level, in both human and mouse placental exchange tissues. We noted that 130 of these genes showed placental phenotypes when

mutated in mice, making these genes good candidates for involvement in human placental pathologies, such as PE and IUGR. Not only were we able to identify new candidates for human disease relevance by direct comparison of expression of orthologs across human and mouse placenta, but we were also able to build out from this core set by network analysis, to identify a novel panel of genes with available mouse knock-out models that may have as yet uncharacterized roles in placenta vascular development. This analysis also revealed two novel members that are ideal candidates for mutational analysis. These genes may also be candidates for patient sample screening as biomarkers with altered expression between normal and IUGR or PE placentas. To streamline this process, unambiguous peptides (i.e. peptides mapping uniquely to only one proteins accession in the IPI database; also termed *proteotypic*) are readily available from our web-page (www. kislingerlab.com) for proteins detected in this study. This will enhance the targeted detection of such candidates by MRM-MS in the future.

Throughout our analysis, we have focused on one-to-one orthologs. However, as we showed for the cathepsin family of proteases, our dataset can be used to detect the similarities and differences in expression between orthologs and paralogs. In particular, of the eight mouse paralogous orthologs for human CTSL, we noted that five of eight still have expression in the mouse placenta as detected by both protein and mRNA similar to our detection of human CTSL in the human villous tree. This expression information will be useful for delineating the evolution of gene expression patterns of paralogous genes.

In summary, we have generated the first detailed proteomic and transcriptomic comparison of a complex tissue of human and mouse. Through tissue microdissection, we were able to enrich for a functionally and physiologically highly related region, the feto-maternal placental exchange tissue. Comprehensive computational analyses revealed high molecular similarity of expressed ortholog gene products and provided a high quality resource for mechanistic investigations using mouse model systems. Protein–protein interaction network analysis has revealed novel candidates likely central to the biology of this tissue. We have generated an intuitive web-interface to the entire data for public access to these data. The entire raw data has been deposited to the Tranche (http://tranche.proteomecommons.org) and GEO (GSE13155; http://www.ncbi.nlm.nih.gov/geo) servers for community access and further bioinformatics data mining.

Materials and methods

Tissue collection

Placentas from naturally mated crosses of C57Bl/6J mice on embryonic day 17.5 were cut into thick slices. A scalpel blade was used to isolate the blood-red labyrinth tissue from the more poorly vascularized, and hence paler, spongiotrophoblast tissue, and to remove the superficial chorionic vasculature. From each litter, 1/4 of the collected labyrinth tissues were set aside for RNA extraction and microarray analysis and 3/4 for cellular fractionation and proteomic analysis, as recently described (Kislinger *et al*, 2006). Human villous trees from two normal placentas delivered by cesarean section at term ($\sim$38 weeks) were dissected and washed in PBS. Tissue was divided for organellar fractionation and RNA extraction as described above (see Supplementary information for details).

Organelle purification and MudPIT analysis

Organelle purification, in-solution digestion and MudPIT analysis were performed as described earlier (Kislinger *et al*, 2006). Individual organelles from pooled tissue samples were analyzed in five replicates on a LTQ linear ion-trap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) using a fully automated 9-cycle, 18 h sequence as described earlier (Washburn *et al*, 2001; Kislinger *et al*, 2006) (see Supplementary information for details). The data associated with this manuscript may be downloaded from ProteomeCommons.org Tranche, http://tranche.proteomecommons.org, using the following hash codes:jhAVbi2rMNNmWPgdD1v5QzvNQwUkMlEn2jSEp3LILFBR wGnMh9jt0SkEOUNJA388o5 + zwlA79P15YOFm2HW7yXIlsd7kAAA AAAADu + Q== and tMrprLnimljkjKeWIqgAIi3JiufeP1BDm + JPbvlb MFnrUzGQTdr5JjowRVkQjIcDbJYwSR8FPJJGS9oMPq/aSWgyjAgAA AAAAADxHQ==

The hash may be used to prove exactly what files were published as part of this manuscript's dataset, and the hash may also be used to check that the data has not changed since publication.

Protein identification

Raw files were converted to m/zXML using ReAdW and searched by X!Tandem against a mouse (v3.28) or human (v3.28) IPI (http:// www.ebi.ac.uk/IPI) protein sequence database. Searches were performed against a target-decoy sequence database and the FDR was set to a conservative level of 0.5% on the peptide level (Gortzak-Uzan *et al*, 2008). Only fully tryptic peptides $\geqslant$7 amino acids, matching these criteria were accepted to generate the final list of identified proteins. We only accepted proteins identified either with two unique peptides or one unique, unambiguous peptide found in $\geqslant$2 MudPIT runs per analyzed organelle fraction. To minimize protein inference, we developed a database grouping scheme and only report proteins with substantial peptide information, as recently reported (Gortzak-Uzan *et al*, 2008; Sodek *et al*, 2008) (see Supplementary information for details).

Comparison to other proteomics datasets

Comparison of our labyrinth proteome resource to our recently published proteomic data were achieved using the ProteinCenter bioinformatics software (Proxeon Biosystems, Odense, Denmark). To link these datasets, we loaded the accession keys for each of the three projects into ProteinCenter and BLASTed them against each other. Only proteins with at least 95% sequence homology were considered identical.

Microarray analysis

Total RNA was extracted from the pooled labyrinths from a litter using Trizol reagent. A total of four litters were used. Balanced aliquots from each litter were then pooled into two technical replicate samples for microarray analysis on Affymetrix MOE430v2 microarrays. Data were processed in GCOS for expression analysis for absent/present calls and signal intensity. Four villi from two individual placentas were separately extracted for total RNA. Two technical replicate samples were prepared by using a balanced aliquot from each of the eight samples. The replicates were analyzed on Affymetrix HG133plus microarray. Data were processed in GCOS for expression analysis for absent/present calls and signal intensity.

Proteomic to microarray comparison

Identified proteins (IPI accessions) and Affymetrix probe annotations were linked to Ensemble gene IDs using the Biomart tool (http:// www.biomart.org) and parsed into a MySQL database and linked on the database level (see Supplementary information for details).

Ortholog comparisons

A table of human-to-mouse gene orthologs was obtained from Ensembl with detailed annotations (i.e. one-to-one, one-to-many and many-to-

many, ortholog scores) and added to the MySQL database. SQL commands were used to link the orthologs and generate a gene ID centered non-redundant list of mouse-to-human orthologs and their corresponding protein and microarray data (see Supplementary information for details).

Jaccard score

A Jaccard score was generated by the comparison of the detection calls generated by GCOS analysis. P, M and A values were transformed into 1 for all P and 0 for M and A. The Jaccard score is the comparison of all attributes (genes) for which there was a measurement. Any attribute for which there was no measurement in either condition is ignored. The Jaccard score, Sij for any two sets (i and j) is given by the equation $S_{ij}=a_{ij}/(a_{ij}+b_{ij}+c_{ij})$, where a, b and c are the sum of all attribute values of (1, 1), (1,0) and (0, 1). In this way, a score of 1 is a perfect correlation and a score of 0 is no correlation.

Detection of phenotypic protein networks

Initially, an interaction network was created consisting of co-expressed orthologous proteins from cluster I and their interactions from the I^{2}D database ver. 1.71. In this network, a set, S, of genes was identified with the phenotypes abnormal placental labyrinth morphology or abnormal placental vasculature. Then genes were identified with statistical significantly enriched connectivity to this set ($P \leqslant 0.05$). A gene had enriched connectivity to S if it had more connections to S than to randomly chosen network subsets, similar to S. To calculate enrichment, 10 000 random network subsets were generated, with the same cardinality and degree distribution as S. Frequencies were then tabulated of a gene with degree, m, having n connections to a random network. These frequencies were used to calculate P-values for the connectivity of genes with S. A set of 34 genes in the network had connectivity P-values $\leqslant 0.05$. A phenotype network was formed from these genes, and their interaction partners from set S.

Note added in proof

GEO has created a new protein/peptide database and has move our deposited data there. The GEO record GSE13299 is still valid for the microarray but the protein data is now this record number PSE115.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Adachi J, Kumar C, Zhang Y, Mann M (2007) In-depth analysis of the adipocyte proteome by mass spectrometry and bioinformatics. *Mol Cell Proteomics* **6:** 1257–1273

Adamson SL, Lu Y, Whiteley KJ, Holmyard D, Hemberger M, Pfarrer C, Cross JC (2002) Interactions between trophoblast cells and the maternal and fetal circulation in the mouse placenta. *Dev Biol* **250:** 358–373

Bandyopadhyay S, Zhan R, Chaudhuri A, Watabe M, Pai SK, Hirota S, Hosobe S, Tsukada T, Miura K, Takano Y, Saito K, Pauza ME, Hayashi S, Wang Y, Mohinta S, Mashimo T, Iiizumi M, Furuta E, Watabe K (2006) Interaction of KAI1 on tumor cells with DARC on vascular endothelium leads to metastasis suppression. *Nat Med* **12:** 933–938

Benfey PN, Mitchell-Olds T (2008) From genotype to phenotype: systems biology meets natural variation. *Science* **320:** 495–497

Bredrup C, Knappskog PM, Majewski J, Rodahl E, Boman H (2005) Congenital stromal dystrophy of the cornea caused by a mutation in the decorin gene. *Invest Ophthalmol Vis Sci* **46:** 420–426

Brown KR, Jurisica I (2007) Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol* **8:** R95

Carter AM (2007) Animal models of human placentation—a review. *Placenta* **28**(Suppl A)**:** S41–S47

Copp AJ (1995) Death before birth: clues from gene knockouts and mutations. *Trends Genet* **11:** 87–93

Cox B, Emili A (2006) Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. *Nat Protoc* **1:** 1872–1878

Cox B, Kislinger T, Emili A (2005) Integrating gene and protein expression data: pattern analysis and profile mining. *Methods* **35:** 303–314

Cox B, Kislinger T, Wigle DA, Kannan A, Brown K, Okubo T, Hogan B, Jurisica I, Frey B, Rossant J, Emili A (2007) Integrated proteomic and transcriptomic profiling of mouse lung development and Nmyc target genes. *Mol Syst Biol* **3:** 109

Deussing J, Kouadio M, Rehman S, Werber I, Schwinde A, Peters C (2002) Identification and characterization of a dense cluster of placenta-specific cysteine peptidase genes and related genes on mouse chromosome 13. *Genomics* **79:** 225–240

Dong JT, Lamb PW, Rinker-Schaeffer CW, Vukanovic J, Ichikawa T, Isaacs JT, Barrett JC (1995) KAI1, a metastasis suppressor gene for prostate cancer on human chromosome 11p11.2. *Science* **268:** 884–886

Egbor M, Ansari T, Morris N, Green CJ, Sibbons PD (2006) Morphometric placental villous and vascular abnormalities in early- and late-onset pre-eclampsia with and without fetal growth restriction. *BJOG* **113:** 580–589

Emes RD, Pocklington AJ, Anderson CN, Bayes A, Collins MO, Vickers CA, Croning MD, Malik BR, Choudhary JS, Armstrong JD, Grant SG (2008) Evolutionary expansion and anatomical specialization of synapse proteome complexity. *Nat Neurosci* **11:** 799–806

Fallon L, Belanger CM, Corera AT, Kontogiannea M, Regan-Klapisz E, Moreau F, Voortman J, Haber M, Rouleau G, Thorarinsdottir T, Brice A, van Bergen En Henegouwen PM, Fon EA (2006) A regulated interaction with the UIM protein Eps15 implicates parkin in EGF receptor trafficking and PI(3)K-Akt signalling. *Nat Cell Biol* **8:** 834–842

Fazioli F, Minichiello L, Matoskova B, Wong WT, Di Fiore PP (1993) eps15, a novel tyrosine kinase substrate, exhibits transforming activity. *Mol Cell Biol* **13:** 5814–5828

Fisher CR, Graves KH, Parlow AF, Simpson ER (1998) Characterization of mice deficient in aromatase (ArKO) because of targeted disruption of the cyp19 gene. *Proc Natl Acad Sci USA* **95:** 6965–6970

Forrai A, Robb L (2005) The gene trap resource: a treasure trove for hemopoiesis research. *Exp Hematol* **33:** 845–856

Foster LJ, de Hoog CL, Zhang Y, Zhang Y, Xie X, Mootha VK, Mann M (2006) A mammalian organelle map by protein correlation profiling. *Cell* **125:** 187–199

Gasperowicz M, Otto F (2008) The notch signalling pathway in the development of the mouse placenta. *Placenta* **29:** 651–659

Ge X, Yamamoto S, Tsutsumi S, Midorikawa Y, Ihara S, Wang SM, Aburatani H (2005) Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics* **86:** 127–141

Gellersen B, Briese J, Oberndorfer M, Redlin K, Samalecos A, Richter DU, Loning T, Schulte HM, Bamberger AM (2007) Expression of the metastasis suppressor KAI1 in decidual cells at the human maternal-fetal interface: regulation and functional implications. *Am J Pathol* **170:** 126–139

Georgiades P, Ferguson-Smith AC, Burton GJ (2002) Comparative developmental anatomy of the murine and human definitive placentae. *Placenta* **23:** 3–19

Gortzak-Uzan L, Ignatchenko A, Evangelou AI, Agochiya M, Brown KA, St Onge P, Kireeva I, Schmitt-Ulms G, Brown TJ, Murphy J, Rosen B, Shaw P, Jurisica I, Kislinger T (2008) A proteome resource of ovarian cancer ascites: integrated proteomic and bioinformatic analyses to identify putative biomarkers. *J Proteome Res* **7:** 339–351

Gramolini AO, Kislinger T, Alikhani-Koopaei R, Fong V, Thompson NJ, Isserlin R, Sharma P, Oudit GY, Trivieri MG, Fagan A, Kannan A, Higgins DG, Huedig H, Hess G, Arab S, Seidman JG, Seidman CE, Frey B, Perry M, Backx PH *et al* (2008) Comparative proteomics profiling of a phospholamban mutant mouse model of dilated cardiomyopathy reveals progressive intracellular stress responses. *Mol Cell Proteomics* **7:** 519–533

Graumann J, Hubner NC, Kim JB, Ko K, Moser M, Kumar C, Cox J, Scholer H, Mann M (2008) Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol Cell Proteomics* **7:** 672–683

Hamada Y, Hiroe T, Suzuki Y, Oda M, Tsujimoto Y, Coleman JR, Tanaka S (2007) Notch2 is required for formation of the placental circulatory system, but not for cell-type specification in the developing mouse placenta. *Differentiation* **75:** 268–278

Hoffmann K, Blaudszun J, Brunken C, Hopker WW, Tauber R, Steinhart H (2005) New application of a subcellular fractionation method to kidney and testis for the determination of conjugated linoleic acid in selected cell organelles of healthy and cancerous human tissues. *Anal Bioanal Chem* **381:** 1138–1144

Kaitu'u-Lino TJ, Morison NB, Salamonsen LA (2007) Estrogen is not essential for full endometrial restoration after breakdown: lessons from a mouse model. *Endocrinology* **148:** 5105–5111

Kim WK, Krumpelman C, Marcotte EM (2008) Inferring mouse gene functions from genomic-scale data using a combined functional network/classification strategy. *Genome Biol* **9**(Suppl 1)**:** S5

Kislinger T, Cox B, Kannan A, Chung C, Hu P, Ignatchenko A, Scott MS, Gramolini AO, Morris Q, Hallett MT, Rossant J, Hughes TR, Frey B, Emili A (2006) Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* **125:** 173–186

Kislinger T, Rahman K, Radulovic D, Cox B, Rossant J, Emili A (2003) PRISM, a generic large scale proteomic investigation strategy for mammals. *Mol Cell Proteomics* **2:** 96–106

Knox K, Baker JC (2008) Genomic evolution of the placenta using co-option and duplication and divergence. *Genome Res* **18:** 695–705

Liao BY, Zhang J (2008) Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc Natl Acad Sci USA* **105:** 6987–6992

Liu H, Sadygov RG, Yates III JR (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76:** 4193–4201

Lu Z, Szafron D, Greiner R, Lu P, Wishart DS, Poulin B, Anvik J, Macdonell C, Eisner R (2004) Predicting subcellular localization of proteins using machine-learned classifiers. *Bioinformatics* **20:** 547–556

Mari G, Hanif F (2007) Intrauterine growth restriction: how to manage and when to deliver. *Clin Obstet Gynecol* **50:** 497–509

Mason RW (2008) Emerging functions of placental cathepsins. *Placenta* **29:** 385–390

Nord AS, Chang PJ, Conklin BR, Cox AV, Harper CA, Hicks GG, Huang CC, Johns SJ, Kawamoto M, Liu S, Meng EC, Morris JH, Rossant J, Ruiz P, Skarnes WC, Soriano P, Stanford WL, Stryke D, von Melchner H, Wurst W *et al* (2006) The International Gene Trap Consortium Website: a portal to all publicly available gene trap cell lines in mouse. *Nucleic Acids Res* **34:** D642–D648

Pasini EM, Kirkegaard M, Salerno D, Mortensen P, Mann M, Thomas AW (2008) Deep coverage mouse red blood cell proteome: a first comparison with the human red blood cell. *Mol Cell Proteomics* **7:** 1317–1330

Pena-Castillo L, Tasan M, Myers CL, Lee H, Joshi T, Zhang C, Guan Y, Leone M, Pagnani A, Kim WK, Krumpelman C, Tian W, Obozinski G, Qi Y, Mostafavi S, Lin GN, Berriz GF, Gibbons FD, Lanckriet G, Qiu J *et al* (2008) A critical assessment of Mus musculus gene function prediction using integrated genomic evidence. *Genome Biol* **9**(Suppl 1)**:** S2

Ponten F, Jirstrom K, Uhlen M (2008) The Human Protein Atlas—a tool for pathology. *J Pathol* **216:** 387–393

Roberts JM, Gammill HS (2005) Preeclampsia: recent insights. *Hypertension* **46:** 1243–1249

Rodriguez TA, Sparrow DB, Scott AN, Withington SL, Preis JI, Michalicek J, Clements M, Tsang TE, Shioda T, Beddington RS, Dunwoodie SL (2004) Cited1 is required in trophoblasts for placental development and for embryo growth and survival. *Mol Cell Biol* **24:** 228–244

Rossant J, Cross JC (2001) Placental development: lessons from mouse mutants. *Nat Rev Genet* **2:** 538–548

Schirmer EC, Florens L, Guan T, Yates III JR, Gerace L (2003) Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science* **301:** 1380–1382

Simpson ER, Zhao Y, Agarwal VR, Michael MD, Bulun SE, Hinshelwood MM, Graham-Lorence S, Sun T, Fisher CR, Qin K, Mendelson CR (1997) Aromatase expression in health and disease. *Recent Prog Horm Res* **52:** 185–213; discussion 213–184

Skarnes WC, von Melchner H, Wurst W, Hicks G, Nord AS, Cox T, Young SG, Ruiz P, Soriano P, Tessier-Lavigne M, Conklin BR, Stanford WL, Rossant J (2004) A public gene trap resource for mouse functional genomics. *Nat Genet* **36:** 543–544

Sodek KL, Evangelou AI, Ignatchenko A, Agochiya M, Brown TJ, Ringuette MJ, Jurisica I, Kislinger T (2008) Identification of pathways associated with invasive behavior by ovarian cancer cells using multidimensional protein identification technology (MudPIT). *Mol Biosyst* **4:** 762–773

Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* **101:** 6062–6067

Sun H, Fang H, Chen T, Perkins R, Tong W (2006) GOFFA: Gene Ontology for Functional Analysis—A FDA Gene Ontology Tool for Analysis of Genomic and Proteomic Data. *BMC Bioinformatics* **7**(Suppl 2)**:** S23

Threadgill DW, Dlugosz AA, Hansen LA, Tennenbaum T, Lichti U, Yee D, LaMantia C, Mourton T, Herrup K, Harris RC, Barnard JA, Yuspa SH, Coffey RJ, Magnuson T (1995) Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science* **269:** 230–234

Washburn MP, Wolters D, Yates III JR (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19:** 242–247

Wolters DA, Washburn MP, Yates III JR (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73:** 5683–5690

Yang JT, Rayburn H, Hynes RO (1993) Embryonic mesodermal defects in alpha 5 integrin-deficient mice. *Development* **119:** 1093–1105

Zhang W, Morris QD, Chang R, Shai O, Bakowski MA, Mitsakakis N, Mohammad N, Robinson MD, Zirngibl R, Somogyi E, Laurin N, Eftekharpour E, Sat E, Grigull J, Pan Q, Peng WT, Krogan N, Greenblatt J, Fehlings M, van der Kooy D *et al* (2004) The functional landscape of mouse gene expression. *J Biol* **3:** 21

Zybailov B, Coleman MK, Florens L, Washburn MP (2005) Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal Chem* **77:** 6218–6224