RNA-Seq reveals conservation of function among the yolk sacs of human, mouse and chicken

Tereza Cindrova-Davies[a], Eric Jauniaux[b], Michael G. Elliot[a,c], Sungsam Gong[d,e], Graham J. Burton[a,*], D. Stephen Charnock-Jones[a,d,e,*]

[a] Centre for Trophoblast Research, Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge, CB2 3EG, UK.
[b] EGA Institute for Women's Health, Faculty of Population Health Sciences. University College London, London, WC1E 6BT, UK.
[c] St John's College, University of Cambridge, Cambridge, CB2 1TP
[d] Department of Obstetrics and Gynaecology, University of Cambridge, The Rosie Hospital, Cambridge, CB2 0SW, UK
[e] National Institute for Health Research, Cambridge Comprehensive Biomedical Research Centre.

* These authors contributed equally

Abbreviated title: Human yolk sac

Corresponding author:
Prof. D. Stephen Charnock-Jones
Department of Obstetrics and Gynaecology,
University of Cambridge
The Rosie Hospital, Robinson Way,
Cambridge CB2 0SW, UK
Tel: +44 1223 336875
Fax: +44 1223 215327
E-mail: dscj1@cam.ac.uk

Abstract

The yolk sac is phylogenetically the oldest of the extraembryonic membranes. The human embryo retains a yolk sac, which goes through primary and secondary phases of development, but its importance is controversial. Although known to synthesize proteins, transport functions are widely considered vestigial. Here, we report RNA-Seq data for the human and murine yolk sacs, and compare with data for the chicken. We also relate the human RNA-Seq data to proteomic data for the coelomic fluid bathing the yolk sac. Conservation of transcriptomes across the species indicates that the human secondary yolk sac likely performs key functions early in development, particularly uptake and processing of macro- and micronutrients, many of which are found in coelomic fluid. More generally, our findings shed light on evolutionary mechanisms that give rise to complex structures such as the placenta. We identify genetic modules that are conserved across mammals and birds, suggesting these are part of the core amniote genetic repertoire and are the building blocks for both oviparous and viviparous reproductive modes. We propose that although a choriovitelline placenta is never established physically in the human, the placental villi, the exocoelomic cavity and the secondary yolk sac function together as a physiological equivalent.

Significance

The human yolk sac is often considered vestigial. Here, we report RNA-Seq analysis of the human and murine yolk sacs, and compare with that of the chicken. We relate the human RNA-Seq data to coelomic fluid proteomic data. Conservation of transcripts across the species indicates the human secondary yolk sac likely performs key functions early in development, particularly uptake and processing of macro- and micronutrients, many of which are found in coelomic fluid. More generally, our findings shed light on evolutionary mechanisms giving rise to complex structures such as the placenta. We propose that although a choriovitelline placenta is never established physically in the human, the placental villi, exocoelomic cavity and secondary yolk sac function together as a physiological equivalent.

\body

Introduction

The yolk sac is phylogenetically the oldest of the extraembryonic membranes, evolving in anamniotes to absorb nutrients from their lipid-rich megalecithal eggs (1). Although the ova of eutherian mammals are microlecithal, the yolk sac has been recruited to transport maternal nutrients during earliest stages of embryonic development. In the majority of species, it makes contact with the chorion to form a transient choriovitelline placenta. This functions during the critical period of organogenesis, at the end of which its functions are generally subsumed by the definitive chorioallantoic placenta. There is, however, considerable species variation, and the most elaborate development is found in rodents and lagomorphs. In these, the yolk sac continues to transport nutrients and immunoglobulins throughout gestation in parallel with the chorioallantoic placenta. For this reason, most of the experimental data on transport have been obtained in the mouse, rat and guinea pig (2-6), and data on the human yolk sac are limited.

The human yolk sac goes through two developmental phases: a primary yolk sac which develops between embryonic days 7 and 9 and is replaced by a secondary yolk sac which is active until day 49 (7). The role of the primary yolk sac is unknown. The importance of the secondary yolk sac remains controversial. Although it is known to synthesize proteins, such as alpha fetoprotein, its transport functions are widely considered vestigial. Primarily, this is because the secondary yolk sac never makes contact with the chorion to form a choriovitelline placenta. Instead, it floats in the exocoelomic cavity, connected to the embryo by the vitelline duct. Although we, and others, have speculated that the yolk sac plays a critical role during organogenesis (3-5, 8-10), there are limited data to support this claim. Obtaining experimental data for the human is impossible for ethical reasons, and thus we adopted an alternative strategy. Here, we report RNA-Seq data derived from human and murine yolk sacs, and compare them with published data from the yolk sac of the chicken. We postulate that conservation of transcripts across these species indicates retention of key transport and synthetic functions. We support this hypothesis by comparing the human yolk sac transcriptome with the proteome of the coelomic fluid.

Results and Discussion

We determined the transcript profile for the first trimester human yolk sac by RNA-Seq with a median sequencing depth of 39 million mapped reads per sample (n=9, SI Appendix Table S1). We identified 12469 transcripts with a mean RPKM (Read Per Kilobase Per Million Mapped Reads) ≥1 (Dataset S1). Similarly, we identified 11628 transcripts in first trimester human placental villous samples (n=11, median sequencing depth 30 million mapped reads, Dataset S2) and 11272 transcripts in the mouse yolk sac (n=8, median sequencing depth 28 million mapped reads, Dataset S3).

In addition, we investigated the protein composition of the coelomic fluid using GELC-MS/MS. We focused on the 165 proteins identified in any 4 of the 5 samples having

excluded immunoglobulins (Dataset S4). Proteins were mapped to unique Ensembl gene identifiers, which were used to identify over-represented GO terms (Dataset S5).

Cholesterol

We selected the 400 most abundant human yolk sac transcripts and identified enriched gene ontology (GO) terms using Panther (complete reference database with Bonferroni correction, Dataset S6). Several terms associated with lipid transport were enriched, for example "very-low-density lipoprotein particle" (23 fold, $P=4.5x10^{-7}$). Indeed, "cholesterol" featured in many of the enriched "biological process" terms (Figure 1). Cholesterol is required for development (3, 5) as it maintains integrity of cell membranes (11), mediates metabolism through propagation of signaling pathways (12), and is the precursor for steroid hormones. In addition, activity of sonic hedgehog (SHH) proteins, which are responsible for the development of the central nervous system (13-15) is determined by covalent modification with cholesterol and other lipids (16). During organogenesis, the embryo is reliant on maternal sources of cholesterol until its liver is sufficiently mature for synthesis (4, 17). Our data show that the human yolk sac contains abundant mRNAs encoding multiple apolipoproteins, the cholesterol efflux transporter ABCA1, as well as lipoprotein receptors, including megalin, cubilin (18), albeit at lower levels (Figure 1). Also present are transcripts encoding all classes of ABC transportors (A to G), which, in addition to transporting cholesterol and lipids, facilitate excretion of toxins and confer multidrug resistance (Table 1). The high abundance (i.e. top 0.5%) of transcripts encoding apolipoproteins present in lipoprotein particles and chylomicrons (ApoB, ApoA1, ApoA2 and ApoA4) is matched by the high levels of these proteins in the coelomic fluid (Dataset S4). Indeed, most of the proteins found in coelomic fluid are highly ranked in the RNA-Seq data (although some were undetectable, i.e. below the RPKM ≥ 1 threshold). Many of these proteins have functions associated with cholesterol or lipid transport and metabolism (Figure 2).

APOB is an essential apolipoprotein found in VLDL particles and chylomicrons. In the latter, which are assembled in enterocytes, the APOB transcript is edited by a cytidine deaminase (APOBEC1, apolipoprotein B mRNA editing enzyme catalytic subunit 1) to introduce a stop codon leading to the production of a short form of the apolipoprotein (APOB-48). The APOB transcripts in the human secondary yolk sac are all of the unedited hepatic form and the APOBEC1 transcript was not detected. Transcripts encoding other proteins required for lipid transport (19) are also abundant, for example, the LDL receptor (LDLR) and microsomal triglyceride transfer protein, MTTP, are in the top 10% and 20 %, respectively. This would suggest that the secondary yolk sac plays a role in facilitating the transport of maternal lipids and cholesterol to the fetal compartments before vascularization of the placental villi is fully established.

The fluid of the exocoelom shares many proteins in common with maternal plasma, supplemented by the addition of specific decidual, trophoblastic and yolk sac proteins. Analogous to maternal serum proteins (20), coelomic fluid proteins can be broadly categorized into common circulating proteins, coagulation and complement factors,

blood transport and binding proteins, protease inhibitors, proteases and other enzymes, cytokines and hormones, channel and receptor-derived peptides and miscellaneous (SI Appendix Table S2) (21, 22).

Transport

The secondary yolk sac comprises an outer mesothelial epithelium and an inner endodermal layer, separated by dilated capillaries and a small amount of mesoderm (23). Both the mesothelial and endodermal epithelia display ultrastructural features typical of an absorptive epithelium, including numerous microvilli, coated pits and pinocytotic vesicles. The GO term "transport" (GO:0006810) was overrepresented in the top 400 yolk sac transcripts (2.87 fold, $p=7.16 \times 10^{-47}$). Many such annotated transcripts were present in the most abundant 20% of transcripts (87) (Dataset S5). Most of these transporter genes are members of the solute carrier (SLC) family of transporters (for example: SLC38A2 (amino acids), SLC4A1 (anions), SLC20A1 (phosphate) and SLC25A37 (iron in the mitochondria)). Transcripts encoding 259 SLC transporters were identified (mean RPKM ranged from 134 to 1). We classified these transporters into 11 main groups on the basis of substrate category (24), i.e. amino acids, urea cycle, glucose, nucleoside sugars, metals, vitamins, neurotransmitters, inorganic ions, thyroid, organic ions, and miscellaneous, and matched them to their main substrates that we and others (21, 22, 25-35) have identified in coelomic fluid samples (Table 2). Zinc is the second most abundant trace element, and is critical for embryonic development. It plays a role in numerous biological processes, including cell division, growth and differentiation and acts as a structural, catalytic and regulatory component within transcription factors, enzymes, transporters and receptors (36, 37). Absorbed zinc is mostly bound to albumin and α2-macroglobulin, both of which are abundant in coelomic fluid. In humans, zinc transport is mediated by 14 members of the ZIP family (SLC39A) and 10 members of the ZnT family (SLC30A). We detected mRNAs encoding 12 ZIP proteins and 8 ZnT proteins in the secondary yolk sac. Immunostaining for the zinc transporter SLC39A7/ZIP7 was present in both the inner endodermal and outer mesothelial epithelia, suggesting uptake from the coelomic fluid and transport to the fetal circulation (Figure 3). The outer mesothelial layer also expresses α-tocopherol transport protein to facilitate vitamin E transport (38). Thus, our data suggest the human secondary yolk sac has a role in the transport of multiple nutrients and vitamins, including iron (seven iron-transporting SLC transcripts identified, Table 2) (39, 40), vitamins A, B12, C, E, and folic acid (Table 2).

Total protein concentration is lower in coelomic fluid than in maternal plasma. However, most amino acids are at higher concentrations and must be derived from the villi and/or the yolk sac. This suggests that the coelomic cavity is an important route for metabolites required for embryonic development (25). The secondary yolk sac floats within this nutrient-rich milieu. It is therefore possible that uterine secretions supplemented by maternal plasma from spiral arteries are taken up by the trophoblast cells, passed via the villous stromal channels into the exocoelomic cavity, from which they are taken up by the yolk sac and transferred to the embryonic gut and the fetal circulation via the vitelline

duct (38). Thus, there appears to be free interchange between these two compartments of the human gestational sac.

The passage across the trophoblast may require lysosomal digestion of macromolecules, and indeed the GO term “lysosome” (GO:0005764) is enriched 4-fold within the most abundant 400 villous transcripts (Bonferroni corrected $P=7.99 \times 10^{-9}$). The efflux amino acid transporters SLC43A2 and SLC7A8 are also highly expressed (above the 95th and 83rd centiles, respectively). In the rat it has been estimated that ~95% of the amino acids provided to the fetus in mid-gestation are obtained by lysosomal digestion of endocytosed maternal proteins (2). In the mouse yolk sac, transcripts encoding 5 lysosomal cathepsins (Ctsl, Ctsz, Ctsb, Ctsd and Ctsh) are among the most abundant 400 transcripts, the activity of which would allow for degradation and release of free amino acids to the developing fetus. We have previously shown that the mesothelial layer of the human yolk sac stains for glycodelin, a product of the uterine glands and present in high concentration in the coelomic fluid (41), indicating exposure to intact maternal proteins and uptake (42). Glycodelin also colocalizes with lysosomal Cathepsin D in human first trimester villi (43). As in the mouse, several cathepsin transcripts (CTSB, CTSZ, CTSL and CTSD) are extremely abundant in the human secondary yolk sac. We found that the GO term “lysosome” is enriched in the list of most abundant transcripts from yolk sacs of human, mouse and chicken (2.79 fold $P=1.35^{-3}$; 3.53 fold $P=9.72^{-5}$ and 3.6 fold $P=2.13^{-3}$, respectively), indicating a similar capacity for digestion of endocytosed macromolecules.

These findings indicate that the exocoelomic cavity is a physiological liquid extension of the early placenta (44), and that the yolk sac is an important route of access for high molecular weight proteins to the embryonic circulation (45).

Hematopoiesis

In all vertebrates, primitive embryonic and definitive fetal/adult blood cells form successively within the yolk sac, fetal liver and bone marrow (46). The human secondary yolk sac is the sole site of hematopoiesis for the first two weeks of pregnancy, and the fetal liver commences blood cell production at week 6 of gestation (47, 48). The term “hemoglobin complex” was significantly enriched among abundant yolk sac transcripts (33 fold, $P=2.95 \times 10^{-6}$). The yolk sac produces predominantly nucleated erythrocytes, which synthesize embryonic hemoglobin (HBZ). There is morphological evidence of the first blood islands in the secondary yolk sac at about day 18 of gestation (49). Yolk sac-derived primitive erythrocytes have been detected in the cardiac cavity as early as the 3-somite stage (21 days), indicative of an established functional network between the yolk sac and embryo (50). The human yolk sac, like that of the mouse, also produces macrophage and multipotential hematopoietic progenitors (48).

Transcription factors

Within the most abundant 400 transcripts in the human yolk sac, 19 genes are annotated as "regulation of transcription, DNA-templated" (GO:0006355), including several

transcription factors (ATF4, FOS, JUN, JUNB and JUND.) In the mouse data set, 8 genes are similarly annotated (SI Appendix Table S3). Candidate motifs recognized by these transcription factors (where known) were identified in the 1kb and 5kb upstream of the TSS of genes which were the highly correlated with the transcription factor transcripts (Datasets S7-10). The FOS and JUN families and AFT4 are closely related, functionally interact, have multiple target genes and are widely expressed. There are numerous candidate binding sites in the highly expressed human yolk sac genes (181 genes with sites for the FOS and JUN families and ATF4, Datasets S7, 8) and 18 genes in the mouse with candidate Atf4 binding sites (Datasets S9, 10). The evidence used for the assignment of GO terms varies and for 2 genes (IGF2 and BHLHE40) depends on a Non-traceable Author Statement (NAS, Dataset S3). Furthermore, binding motifs are not available for all candidate factors even in the most recent JASPAR database.

Yolk sac and villi compared to adult lung, liver and kidney

The inaccessibility of the human yolk sac severely constrains any functional investigation of these candidate transcription binding sites. We therefore compared the human yolk sac transcript profile with tissues where function has been better defined experimentally. The placental villi serve similar functions to the adult lung, liver and kidney, and we therefore compared the overlap among the 400 most abundant transcripts from these tissues and the yolk sac (Figure 4). The transcripts that are unique to each tissue and those shared among these tissues are listed in Dataset S11. As expected, the transcripts shared by all 5 tissues encoded abundant house-keeping proteins, such as ribosomal proteins and those involved in mitochondrial energy generation with GO terms such as “cytosolic small ribosome subunit”, “cytosolic large ribosome subunit” and “mitochondrial respiratory chain” being significantly overrepresented ($P< 8.5x10^{-8}$ after Bonferroni correction, Dataset S12). The enriched GO terms associated with the 35 transcripts shared only by the yolk sac and liver include “high-density lipoprotein particle receptor binding”, “cholesterol transporter activity”, “lipid transporter activity” (all greater than 28-fold enrichment and $P< 2x10^{-4}$ after Bonferroni correction, Dataset S13).

We also examined the overlap among over-represented GO terms in the most abundant transcripts in these 5 tissues (Supplementary Datasets S6, 14-17). For example, the Biological Process terms "high-density lipoprotein particle remodeling", "lipid transport", "cholesterol efflux", and "cholesterol metabolic process" are all overrepresented and shared between the yolk sac and liver (Figure 5, Datasets S18-20). It is striking that among the shared transcripts many (18 of 35) encode proteins with transport capacity. Besides the lipid and cholesterol transport-associated genes (APOA1, APOA2, APOA4, APOB, APOC1, APOC3 and APOH), the major serum carriers of corticosteroids and progesterone (SERPINA6) are present. In addition, the thyroxine (T4) hormone carrier (SERPINA7 also known as TBG, and TTR) was detected. Maternal thyroxine is essential for fetal brain development prior to the commencement of fetal thyroid function at the end of the first trimester (26). Furthermore, transcripts encoding several metal binding or transporting proteins are present (MT1E, MT1H, TF, HPX and SEPP1). The latter is a selenoprotein which contains 10 selenocysteine residues per molecule, and is

postulated to function as a carrier of selenium (51). These data indicate that the yolk sac shares many abundant transcripts with other metabolically active fetal and adult tissues, and that it contains transcripts that encode several liver specific functions, including those encoding several essential hormone transporters.

Mouse Yolk Sac RNA-Seq data

On a comparative basis, the yolk sac provides an important pathway for nutrient uptake in many species during early pregnancy (52, 53), and its role has been well documented in experimental rodents, such as the rat and mouse (54). At the ultrastructural level, the endodermal layers of the human and rodent yolk sacs display many similar morphological characteristic typical of an absorptive epithelium (10, 55). However, their orientation is different as the human yolk sac floats in the exocoelom with the endodermal cell layer lining the interior of the sac, whilst the rodent yolk is inverted with the endodermal layer facing outwards following degeneration of the parietal layer (56, 57).

Cross-species comparison

To compare the most abundant transcripts in the human, mouse and chicken yolk sac we identified the homologous genes by mapping the human Ensembl gene identifiers to the corresponding mouse and chicken Ensembl gene IDs using biomart (release ENSEMBL Genes 85) Dataset S20. The raw and processed chicken yolk sac data were obtained directly from the authors (58). Gene ontology analysis was carried out as described above, using the 400 most abundant chicken yolk sac transcripts observed at E17 (Dataset S22). The intersection between the over-represented GO terms in the human, mouse and chicken yolk sacs was determined as described above (SI Appendix Table S4). Venn diagrams showing the overlaps are shown in Figure 6. The P values for all the observed intersections are highly significant ($P< 1.70x10^{-22}$ and are summarized in Dataset S23). The depth of the GO annotation varies between species and is more limited in the chicken, which necessarily precludes detecting a high degree of overlap with the human genes. Nonetheless, shared biological process terms include “translation”, “ribosome biogenesis”, “oxidation-reduction process” and “small molecule metabolic process” reflecting cellular processes associated with metabolically active tissue. Transcripts reflecting more specialized cellular function, such as lipid and cholesterol transport, which are shared by the human yolk sac and liver (APOA1, APOA4, APOB, RBP4, SEPP1, TTR) are also present among the 400 most abundant mouse and chicken yolk sac transcripts ($P=2.3 x10^{-13}$).

Conclusion

Overall, our data show that the human secondary yolk sac appears far from vestigial with little or no biological role, and that it may perform many key functions in the early weeks of development. In particular, it contains abundant transcripts encoding an array of transporter proteins involved in the uptake of macro- and micronutrients. Our data showing the presence of transporter proteins in both of the yolk sac epithelia, and their substrates in coelomic fluid, support this concept and are consistent with the

morphological appearances of absorptive surfaces (23, 59, 60). Our data also confirm significant synthetic activity, especially of apolipoproteins, a function most likely performed in the endodermal cells given their high content of endoplasmic reticulum and Golgi cisternae (23, 59, 60). The handling of cholesterol, which is essential for synthesis of cell and organelle membranes as well as being a co-factor in signaling pathways involved in axis determination and other fundamental developmental events, appears of particular significance. Furthermore, the high level of conservation of transcripts compared with the mouse and the chicken, where the yolk sac is known to be essential for development, suggests maintenance of function.

The secondary yolk sac is thus likely to be essential for the survival of the embryo during the first weeks of development. Morphological abnormalities of the secondary yolk sac have been reported in 70% of first trimester human miscarriages (61), but separating cause from effect is impossible.

More generally, our findings shed light on the evolutionary mechanisms that give rise to complex structures such as the placenta. The placenta has evolved repeatedly from an oviparous background in mammals, reptiles and fish, sometimes over very short timescales (62, 63). Development of a yolk sac for direct maternal provisioning of the developing fetus is a common feature of such evolutionary transitions to the extent that the yolk sac may be regarded as a "fundamental vertebrate fetal nutritional system" (64). The yolk sacs studied in this paper constitute a broad, albeit incomplete, sampling of the variation present in the vertebrates. The inverted yolk sac placenta of the mouse is found in several rodents and lagomorphs (1), but also in the distantly related nine-banded armadillo (65). The "free floating" secondary yolk sac of the human is found in the other haplorhine primates (*Macaca mulatta*) (66) but also, surprisingly, in distantly related Afrotherian species. The yolk sac floats freely in the exocoelom in at least one tenrec species, the Nimba otter shrew (*Micropotamogale lamottei*) (67), in another Afrotherian insectivore, the eastern rock elephant shrew (*Elephantulus myurus*) (68) and in two species of bat (*Myotis lucifugus* (little brown bat) (69) and *Tadarida brasiliensis* (70)). These observations suggest that both mouse and human yolk sacs reflect convergent evolution toward similar forms found elsewhere in the mammalian phylogenetic tree. The yolk sac of the chicken, of course, is characteristic of oviparous species: all birds and the majority of reptiles.

Under a classic Darwinian model of macroevolution, small genetic mutations gradually occur and are accompanied by correspondingly small phenotypic changes until, over time, a high degree of morphological and functional diversification accumulates between species that are distantly related. The results presented here suggest, however, that the genetic systems underlying the function of the yolk sac are robust. We identify a number of genetic modules, including those involved in cholesterol processing, lipid transport, redox processes and nutrient delivery which were presumably reorganized and redeployed during evolutionary change while being internally conserved. These findings are in line with an extended evolutionary-developmental model in which it is the gene

regulatory networks and underlying transcriptional control elements which change (71). The repeated convergent evolution of yolk sac placentas in all major groups of vertebrates other than birds is characteristic of what has come to be known as *deep homology* which describes the origin of complex structures through modification and reorganization of pre-existing genetic systems (72). Given that the modular conservation of systems active in the yolk sac is shown to extend across mammals and birds, the common ancestor of which was a reptile, it is possible that genetic modules of the yolk sac are part of the *core amniote genetic repertoire*. That is, conserved genetic systems of yolk sac function, for example cholesterol and lipid metabolism, form part of the common heritage shared by all mammals, reptiles and birds, and are the building blocks for both oviparous and viviparous reproductive modes.

Our findings indicate that extensive high level morphological diversification of the extraembryonic membranes masks a surprising degree of functional conservation at the molecular genetic level. Evolutionary conservation at the level of nucleotide sequence, gene regulation and modularity of gene expression is widely regarded as evidence of functional significance in both healthy development and in disease (73-75). Therefore, we propose, that although a choriovitelline placenta is never established physically in the human, the early placental villi, the exocoelomic cavity and the secondary yolk sac combine to function as a physiological equivalent (Figure 7).

Methods

Human tissue collection

Tissue and fluid samples were collected with informed written patient consent and approval of the Joint UCL/UCLH Committees on the Ethics of Human Research (05/Q0505/82) from 7-12 weeks in uncomplicated pregnancies. Gestational age was confirmed by ultrasound measurement of the crown rump length of the embryo. All samples were collected from patients undergoing surgical pregnancy termination under general anesthesia for psycho-social reasons. Coelomic fluid samples were obtained by transvaginal puncture under sonographic guidance as previously described (27). Villous samples were obtained under transabdominal ultrasound guidance from the central region of the placenta using a chorionic villus sampling (CVS) technique. Intact secondary yolk sacs were obtained by gentle aspiration guided by ultrasound. All samples were snap frozen in liquid nitrogen and stored at -80°C until analysis.

Mouse tissue collection

Yolk sacs were collected from time-mated virgin C57BL/6J mice. Experiments were carried out in accordance with the United Kingdom Animals Scientific Procedures Act 1986 which mandates ethical review. A single randomly selected yolk sac was collected from each pregnant female at E9.5 (day of plug = E0.5). Tissue was dissected free from decidua and amnion, snap frozen and stored at -80°C until processing.

RNA extraction and RNA-Seq

RNA from human and mouse yolk sacs and human first trimester placental villi was extracted using RNeasy plus universal mini kit (Qiagen, Cat No 73404). Libraries were made using the Illumina TruSeq Stranded mRNA library kit according to the manufacturer's instructions. Libraries were quantified (kappa qPCR) and equimolar pools sequenced (single end 50 base reads, SE50) in several lanes of the Illumina HiSeq2500. Additional details are provided in the Supplemental Information.

Data availability

The data sets generated during the current study are available in the European Nucleotide Archive (ENA http://www.ebi.ac.uk/ena) under the accession number PRJEB18767, http://www.ebi.ac.uk/ena/data/view/PRJEB18767.

Proteomic analysis of coelomic fluid samples

Coelomic fluid samples were run on 1D gels, enzymatically digested and analyzed using LC-MS/MS (Dionex Ultimate 3000 RSLC nanoUPLC, Thermo Fisher Scientific Inc, Waltham, MA, USA) system and a QExactive Orbitrap mass spectrometer (Thermo Fisher Scientific Inc, Waltham, MA, USA). Detailed methods are described in the Supplemental file.

Immunohistochemistry

Immunohistochemistry was performed as previously described (76) using the following primary antibodies: anti-SLC39A7 (ZIP7, Abcam, ab117560) and anti-ABCA1 (Abcam, ab7360).

Supporting Information.

Additional methods and references (57, 77-87) are in the online supplemental methods.

Acknowledgements

This study was supported by the Medical Research Council (MR/L020041/1). We would like to thank Prof Kathryn Lilley (University of Cambridge) for the proteomic analysis and Dr. Erica Watson (University of Cambridge) for assistance with the mouse yolk sac samples. M.G.E. is the recipient of a Research Fellowship from St John's College, Cambridge.

Author contributions.

T.C-D., E.J., G.J.B. and D.S.C.-J. designed the study.
T.C.-D., E.J. and S.G. performed the experiments.
T.C.-D., G.J.B. and D.S.C.-J. wrote the manuscript with contributions from E.J., M.G.E. and S.G.
All authors analyzed data.
All authors edited and approved the final manuscript

Competing financial interests

The authors declare no competing financial interests.

Corresponding author:
Prof. D. Stephen Charnock-Jones

References

1. Mossman, H. (1987) *Vertebrate fetal membranes: comparative ontogeny and morphology: evolution; phylogenetic significance; basic finctions, research opportunities* (Macmillan, London).
2. Brent, R. L. & Fawcett, L. B. (1998) Nutritional studies of the embryo during early organogenesis with normal embryos and embryos exhibiting yolk sac dysfunction. *J Pediatr* **132:**S6-16.
3. Woollett, L. A. (2008) Where does fetal and embryonic cholesterol originate and what does it do? *Annu Rev Nutr* **28:**97-114.
4. Baardman, M. E., Erwich, J. J., Berger, R. M., Hofstra, R. M., Kerstjens-Frederikse, W. S.*, et al.* (2012) The origin of fetal sterols in second-trimester amniotic fluid: endogenous synthesis or maternal-fetal transport? *Am J Obstet Gynecol* **207:**202 e219-225.
5. Baardman, M. E., Kerstjens-Frederikse, W. S., Berger, R. M., Bakker, M. K., Hofstra, R. M.*, et al.* (2013) The role of maternal-fetal cholesterol transport in early fetal life: current insights. *Biol Reprod* **88:**24.
6. King, B. F. & Enders, A. C. (1970) Protein absorption and transport by the guinea pig visceral yolk sac placenta. *Am J Anat* **129:**261-287.
7. Moore, K. (1988) *The Developing Human Clinically Oriented Embryology* (WB Saunders Company, Harcourt Brace Jovanovich, Inc., Philadelphia, London, Toronto, Montreal, Sydney, Tokyo).
8. Zohn, I. E. & Sarkar, A. A. (2010) The visceral yolk sac endoderm provides for absorption of nutrients to the embryo during neurulation. *Birth Defects Res A Clin Mol Teratol* **88:**593-600.
9. Jauniaux, E., Gulbis, B., & Burton, G. J. (2003) The human first trimester gestational sac limits rather than facilitates oxygen transfer to the foetus--a review. *Placenta* **24 Suppl A:**S86-93.
10. Jones, C. (1997) in *Embryonic medicine and therapy*, eds. Jaunaiux, E., Barnea, E., & Edwards, R. (Oxford University Press, Oxford).
11. Porn, M. I., Ares, M. P., & Slotte, J. P. (1993) Degradation of plasma membrane phosphatidylcholine appears not to affect the cellular cholesterol distribution. *J Lipid Res* **34:**1385-1392.
12. Fielding, C. J. & Fielding, P. E. (2004) Membrane cholesterol and the regulation of signal transduction. *Biochem Soc Trans* **32:**65-69.
13. Porter, J. A., Young, K. E., & Beachy, P. A. (1996) Cholesterol modification of hedgehog signaling proteins in animal development. *Science* **274:**255-259.
14. Cooper, M. K., Wassif, C. A., Krakowiak, P. A., Taipale, J., Gong, R.*, et al.* (2003) A defective response to Hedgehog signaling in disorders of cholesterol biosynthesis. *Nat Genet* **33:**508-513.
15. Marti, E. & Bovolenta, P. (2002) Sonic hedgehog in CNS development: one signal, multiple outputs. *Trends Neurosci* **25:**89-96.
16. Long, J., Tokhunts, R., Old, W. M., Houel, S., Rodgriguez-Blanco, J.*, et al.* (2015) Identification of a family of fatty-acid-speciated sonic hedgehog proteins, whose members display differential biological properties. *Cell Rep* **10:**1280-1287.
17. Witsch-Baumgartner, M., Gruber, M., Kraft, H. G., Rossi, M., Clayton, P.*, et al.* (2004) Maternal apo E genotype is a modifier of the Smith-Lemli-Opitz syndrome. *J Med Genet* **41:**577-584.
18. Burke, K. A., Jauniaux, E., Burton, G. J., & Cindrova-Davies, T. (2013) Expression and immunolocalisation of the endocytic receptors megalin and cubilin in the human yolk sac and placenta across gestation. *Placenta* **34:**1105-1109.

19. Xiao, C., Hsieh, J., Adeli, K., & Lewis, G. F. (2011) Gut-liver interaction in triglyceride-rich lipoprotein metabolism. *Am J Physiol Endocrinol Metab* **301:**E429-446.
20. Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J., Angell, N. H.*, et al.* (2002) Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. *Mol Cell Proteomics* **1:**947-955.
21. Jauniaux, E., Gulbis, B., Jurkovic, D., Campbell, S., Collins, W. P.*, et al.* (1994) Relationship between protein concentrations in embryological fluids and maternal serum and yolk sac size during human early pregnancy. *Hum Reprod* **9:**161-166.
22. Jauniaux, E. & Gulbis, B. (2000) Fluid compartments of the embryonic environment. *Human Reproduction Update* **6:**268-278.
23. Jones, C. J. P. & Jauniaux, E. (1995) Ultrastructure of the materno-embryonic interface in the first trimester of pregnancy. *Micron* **26:**145-173.
24. Lin, L., Yee, S. W., Kim, R. B., & Giacomini, K. M. (2015) SLC transporters as therapeutic targets: emerging opportunities. *Nat Rev Drug Discov* **14:**543-560.
25. Jauniaux, E., Gulbis, B., Gerlo, E., & Rodeck, C. (1998) Free amino acid distribution inside the first trimester human gestational sac. *Early Hum Dev* **51:**159-169.
26. Contempre, B., Jauniaux, E., Calvo, R., Jurkovic, D., Campbell, S.*, et al.* (1993) Detection of thyroid hormones in human embryonic cavities during the first trimester of pregnancy. *J Clin Endocrinol Metab* **77:**1719-1722.
27. Jauniaux, E., Sherwood, R. A., Jurkovic, D., Boa, F. G., & Campbell, S. (1994) Amino acid concentrations in human embryological fluids. *Hum Reprod* **9:**1175-1179.
28. Campbell, J., Wathen, N., Macintosh, M., Cass, P., Chard, T.*, et al.* (1992) Biochemical composition of amniotic fluid and extraembryonic coelomic fluid in the first trimester of pregnancy. *Br J Obstet Gynaecol* **99:**563-565.
29. Jauniaux, E., Jurkovic, D., Gulbis, B., Gervy, C., Ooms, H. A.*, et al.* (1991) Biochemical composition of exocoelomic fluid in early human pregnancy. *Obstet Gynecol* **78:**1124-1128.
30. Jauniaux, E., Hempstock, J., Teng, C., Battaglia, F. C., & Burton, G. J. (2005) Polyol concentrations in the fluid compartments of the human conceptus during the first trimester of pregnancy: maintenance of redox potential in a low oxygen environment. *J Clin Endocrinol Metab* **90:**1171-1175.
31. Gulbis, B., Jauniaux, E., Decuyper, J., Thiry, P., Jurkovic, D.*, et al.* (1994) Distribution of iron and iron-binding proteins in first-trimester human pregnancies. *Obstet Gynecol* **84:**289-293.
32. Wathen, N. C., Delves, H. T., Campbell, D. J., & Chard, T. (1995) The coelomic cavity--a reservoir for metals. *Am J Obstet Gynecol* **173:**1884-1888.
33. Campbell, J., Wathen, N., Perry, G., Soneji, S., Sourial, N.*, et al.* (1993) The coelomic cavity: an important site of materno-fetal nutrient exchange in the first trimester of pregnancy. *Br J Obstet Gynaecol* **100:**765-767.
34. Campbell, J., Wathen, N. C., Merryweather, I., Abbott, R., Muller, D.*, et al.* (1994) Concentrations of vitamins A and E in amniotic fluid, extraembryonic coelomic fluid, and maternal serum in the first trimester of pregnancy. *Arch Dis Child Fetal Neonatal Ed* **71:**F49-50.
35. Iles, R. K., Wathen, N. C., Sharma, K. B., Campbell, J., Grudzinskas, J. G.*, et al.* (1994) Pregnancy-associated plasma protein A levels in maternal serum, extraembryonic coelomic and amniotic fluids in the first trimester. *Placenta* **15:**693-699.

36. Vallee, B. L. & Falchuk, K. H. (1993) The biochemical basis of zinc physiology. *Physiol Rev* **73:**79-118.
37. Kambe, T., Hashimoto, A., & Fujimoto, S. (2014) Current understanding of ZIP and ZnT zinc transporters in human health and diseases. *Cell Mol Life Sci* **71:**3281-3295.
38. Jauniaux, E., Cindrova-Davies, T., Johns, J., Dunster, C., Hempstock, J.*, et al.* (2004) Distribution and transfer pathways of antioxidant molecules inside the first trimester human gestational sac. *J Clin Endocrinol Metab* **89:**1452-1459.
39. Gulbis, B., Jauniaux, E., Cotton, F., & Stordeur, P. (1998) Protein and enzyme patterns in the fluid cavities of the first trimester gestational sac: relevance to the absorptive role of secondary yolk sac. *Mol Hum Reprod* **4:**857-862.
40. Evans, P., Cindrova-Davies, T., Muttukrishna, S., Burton, G. J., Porter, J.*, et al.* (2011) Hepcidin and iron species distribution inside the first-trimester human gestational sac. *Mol Hum Reprod* **17:**227-232.
41. Wathen, N. C., Cass, P. L., Campbell, D. J., Kitau, M. J., & Chard, T. (1992) Levels of placental protein 14, human placental lactogen and unconjugated oestriol in extraembryonic coelomic fluid. *Placenta* **13:**195-197.
42. Burton, G. J., Watson, A. L., Hempstock, J., Skepper, J. N., & Jauniaux, E. (2002) Uterine glands provide histiotrophic nutrition for the human fetus during the first trimester of pregnancy. *J Clin Endocrinol Metab* **87:**2954-2959.
43. Hempstock, J., Cindrova-Davies, T., Jauniaux, E., & Burton, G. J. (2004) Endometrial glands as a source of nutrients, growth factors and cytokines during the first trimester of human pregnancy: a morphological and immunohistochemical study. *Reprod Biol Endocrinol* **2:**58.
44. Jauniaux, E., Gulbis, B., Hyett, J., & Nicolaides, K. H. (1998) Biochemical analyses of mesenchymal fluid in early pregnancy. *Am J Obstet Gynecol* **178:**765-769.
45. Castellucci, M. & Kaufmann, P. (1982) A three-dimensional study of the normal human placental villous core: II. Stromal architecture. *Placenta* **3:**269-285.
46. Baron, M. H., Vacaru, A., & Nieves, J. (2013) Erythroid development in the mammalian embryo. *Blood Cells Mol Dis* **51:**213-219.
47. Tavian, M. & Peault, B. (2005) The changing cellular environments of hematopoiesis in human development in utero. *Exp Hematol* **33:**1062-1069.
48. Palis, J. & Yoder, M. C. (2001) Yolk-sac hematopoiesis: the first blood cells of mouse and man. *Exp Hematol* **29:**927-936.
49. Luckett, W. P. (1978) Origin and differentiation of the yolk sac and extraembryonic mesoderm in presomite human and rhesus monkey embryos. *Am J Anat* **152:**59-97.
50. Tavian, M., Hallais, M. F., & Peault, B. (1999) Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. *Development* **126:**793-803.
51. Burk, R. F. & Hill, K. E. (2005) Selenoprotein P: an extracellular protein with unique physical characteristics and a role in selenium homeostasis. *Annu Rev Nutr* **25:**215-235.
52. Wooding, F. & Flint, A. (1994) in *Marshall's Physiology of Reproduction*, ed. Lamming, G. (Chapman & Hall, London).
53. Wooding, F. P. & Burton, G. J. (2008) *Comparative Placentation. Structures, Functions and Evolution* (Springer, Berlin).
54. Lloyd, J. B., Beckman, D. A., & Brent, R. L. (1998) Nutritional role of the visceral yolk sac in organogenesis-stage rat embryos. *Reprod Toxicol* **12:**193-195.

55. Carter, A. M. (2016) IFPA Senior Award Lecture: Mammalian fetal membranes. *Placenta* **48 Suppl 1:**S21-S30.
56. Burton, G. J., Hempstock, J., & Jauniaux, E. (2001) Nutrition of the human fetus during the first trimester--a review. *Placenta* **22 Suppl A:**S70-77.
57. Pereda, J., Correr, S., & Motta, P. M. (1994) The structure of the human yolk sac: a scanning and transmission electron microscopic analysis. *Arch Histol Cytol* **57:**107-117.
58. Yadgary, L., Wong, E. A., & Uni, Z. (2014) Temporal transcriptome analysis of the chicken embryo yolk sac. *BMC Genomics* **15:**690.
59. Hoyes, A. D. (1969) The human foetal yolk sac. An ultrastructural study offour specimens. *Z Zellforsch Mikrosk Anat* **99:**469-490.
60. Hesseldahl, H. & Larsen, J. F. (1969) Ultrastructure of human yolk sac: endoderm, mesenchyme, tubules and mesothelium. *Am J Anat* **126:**315-335.
61. Nogales, F., Beltran, E., & Gonzales, F. (1993) in *The human yolk sac and yolk sac tumors* (Springer, Berlin, Heidelberg).
62. Blackburn, D. G. (1992) Convergent Evolution of Viviparity, Matrotrophy, and Specializations for Fetal Nutrition in Reptiles and Other Vertebrates. *Amer. Zool.* **32:**313-321.
63. Dulvy, N. & Reynolds, J. (1997) Evolutionary transition among egg-laying, live-bearing and maternal inputs in sharks and rays. *Proceedings of the Royal Society B: Biological Sciences* **264(1386):**1309-1315.
64. Stewart, J. (1993) Yolk Sac Placentation in Reptiles: Structural Innovation in a Fundamental Vertebrate Fetal Nutritional System. *J Exp Zool* **266:**431-449.
65. Enders, A. C. (1960) Development and structure of the villous haemochorial placenta of the nine-banded armadillo (Dasypus novemcinctus). *J Anat* **94:**34-45.
66. King, B. F. & Wilson, J. M. (1983) A fine structural and cytochemical study of the rhesus monkey yolk sac: endoderm and mesothelium. *Anat Rec* **205:**143-158.
67. Carter, A. M., Blankenship, T. N., Enders, A. C., & Vogel, P. (2006) The fetal membranes of the otter shrews and a synapomorphy for afrotheria. *Placenta* **27:**258-268.
68. Oduor-Okelo, D., Katema, R. M., & Carter, A. M. (2004) Placenta and fetal membranes of the four-toed elephant shrew, Petrodromus tetradactylus. *Placenta* **25:**803-809.
69. Enders, A. C., Wimsatt, W. A., & King, B. F. (1976) Cytological development of yolk sac endoderm and protein-absorptive mesothelium in the little brown bat, Myotis lucifugus. *Am J Anat* **146:**1-30.
70. Stephens, R. J. & Cabral, L. J. (1971) Cytological differentiation of the mesothelial cells of the yolk sac of the bat, Tadarida brasiliensis cynocephala. *Anat Rec* **171:**293-312.
71. Carroll, S. B. (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134:**25-36.
72. Shubin, N., Tabin, C., & Carroll, S. (2009) Deep homology and the origins of evolutionary novelty. *Nature* **457:**818-823.
73. Boffelli, D., Nobrega, M. A., & Rubin, E. M. (2004) Comparative genomics at the vertebrate extremes. *Nat Rev Genet* **5:**456-465.
74. Stewart, J. B., Freyer, C., Elson, J. L., & Larsson, N. G. (2008) Purifying selection of mtDNA and its implications for understanding evolution and mitochondrial disease. *Nat Rev Genet* **9:**657-662.
75. Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K.*, et al.* (2005) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* **3:**e7.

76. Cindrova-Davies, T., Yung, H. W., Johns, J., Spasic-Boskovic, O., Korolchuk, S., *et al.* (2007) Oxidative Stress, Gene Expression, and Protein Changes Induced in the Human Placenta during Labor. *Am J Pathol* **171:**1168-1179.
77. Martin, M. (2016) Cutadapt removes adapter sequences from high-throughput sequencing reads. doi 10.14806/ej.17.1.200.
78. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., *et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14:**R36.
79. Anders, S., Pyl, P. T., & Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31:**166-169.
80. Love, M. I., Huber, W., & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:**550.
81. Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., *et al.* (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* **13:**397-406.
82. Mi, H., Muruganujan, A., & Thomas, P. D. (2013) PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* **41:**D377-386.
83. Walter, W., Sanchez-Cabo, F., & Ricote, M. (2015) GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics* **31:**2912-2914.
84. Wang, M., Zhao, Y., & Zhang, B. (2015) Efficient Test and Visualization of Multi-Set Intersections. *Sci Rep* **5:**16923.
85. Benjamini, Y. & Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J Royal Stat Society Series B* **57(1):**289-300.
86. Mele, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., *et al.* (2015) Human genomics. The human transcriptome across tissues and individuals. *Science* **348:**660-665.
87. Jauniaux, E., Jurkovic, D., Henriet, Y., Rodesch, F., & Hustin, J. (1991) Development of the secondary human yolk sac: correlation of sonographic and anatomical features. *Hum Reprod* **6:**1160-1166.

Figure Legends

Figure 1 Chord plot illustrating the GO biological process terms that include “cholesterol” and that are overrepresented in the 400 most abundant yolk sac transcripts on the right, and the genes contributing to that enrichment on the left arranged in order of their expression level.

Figure 2 Chord plot illustrating proteins present in the coelomic fluid that relate to GO biological process terms involving “cholesterol” and “lipid transport”. The presence of these proteins is consistent with the high level of their transcripts in the human yolk sac illustrated in Figure 1.

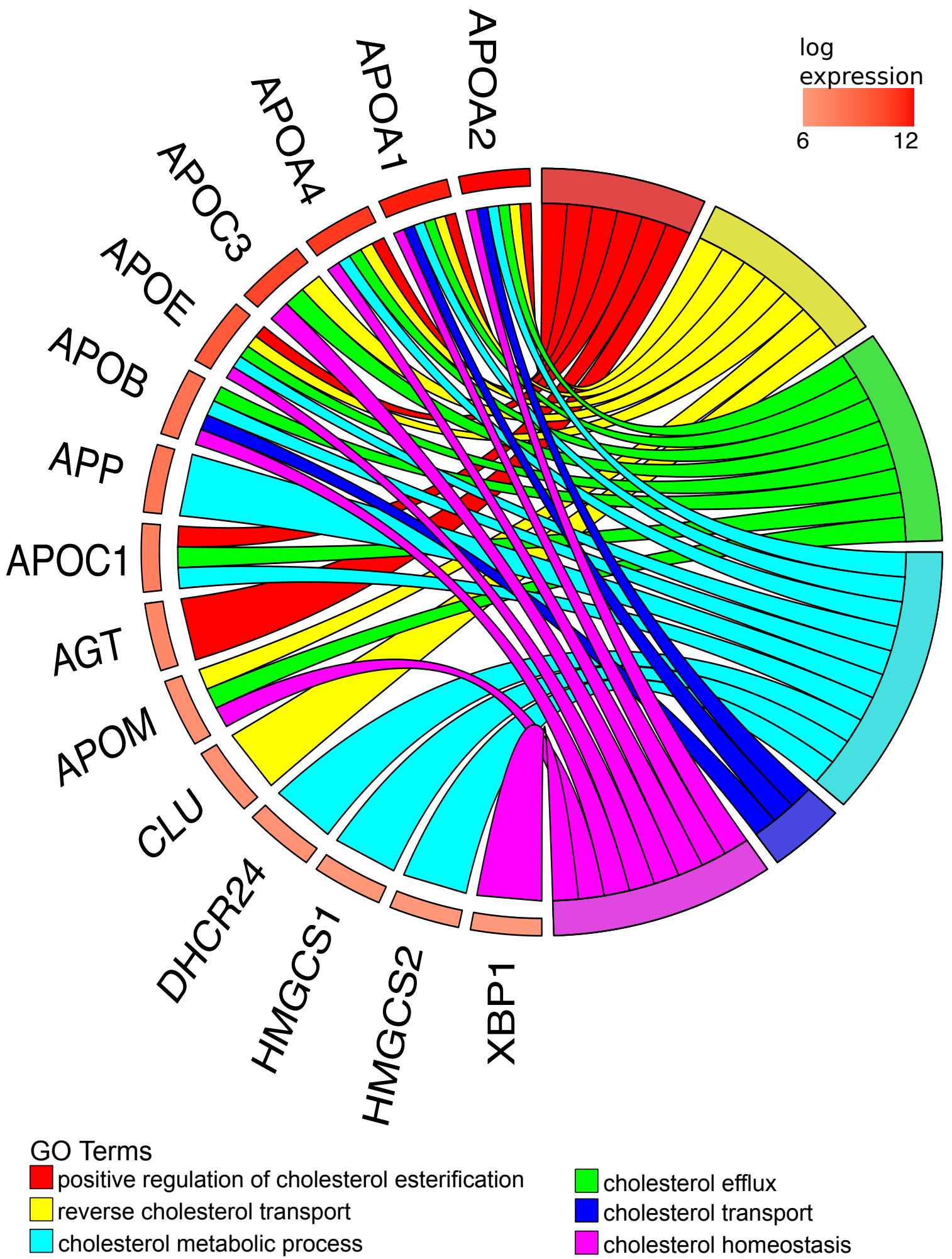
Figure 3 Immunolocalization of ABCA1 and SLC39A7/ZIP7 transporter proteins in the human yolk sac at 11 weeks gestational age. Sections were immunostained with anti-ABCA1 or anti-ZIP7 antibodies. In both cases, staining was present in the inner endodermal and outer mesothelial layers, although it was stronger in the former.

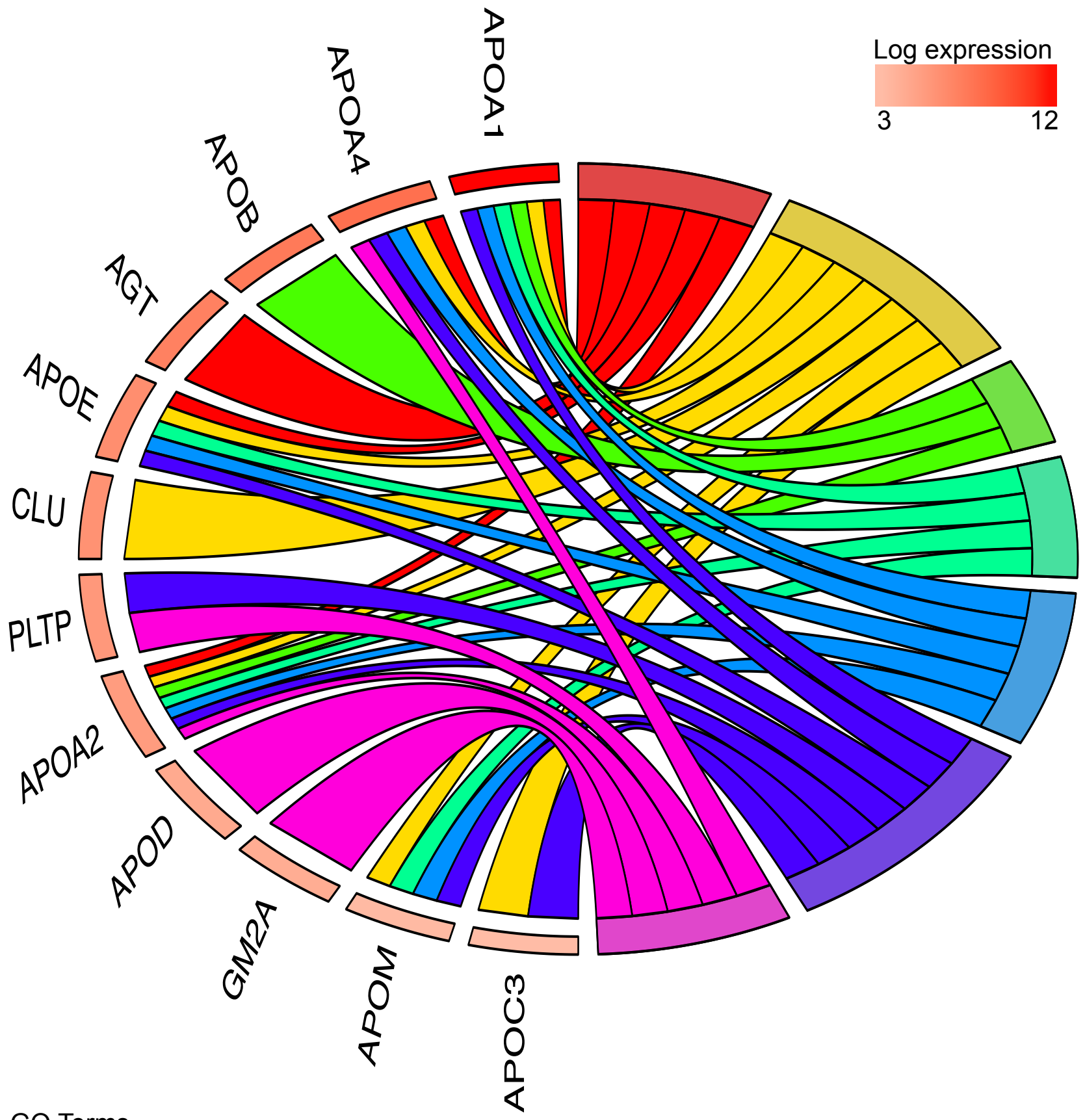
Figure 4 Venn diagram comparing the most abundant 400 transcripts in the human yolk sac, with first trimester placental villi and adult liver, lung and kidney. Transcripts shared by all five tissues (83) encoded principally housekeeping proteins, whereas those shared uniquely with liver (35) encoded proteins involved in cholesterol and lipid metabolism, suggesting the yolk sac may perform these functions while the fetal liver develops. By contrast, there are few transcripts shared uniquely with the kidney (5) suggesting the yolk sac plays little role in excretion.

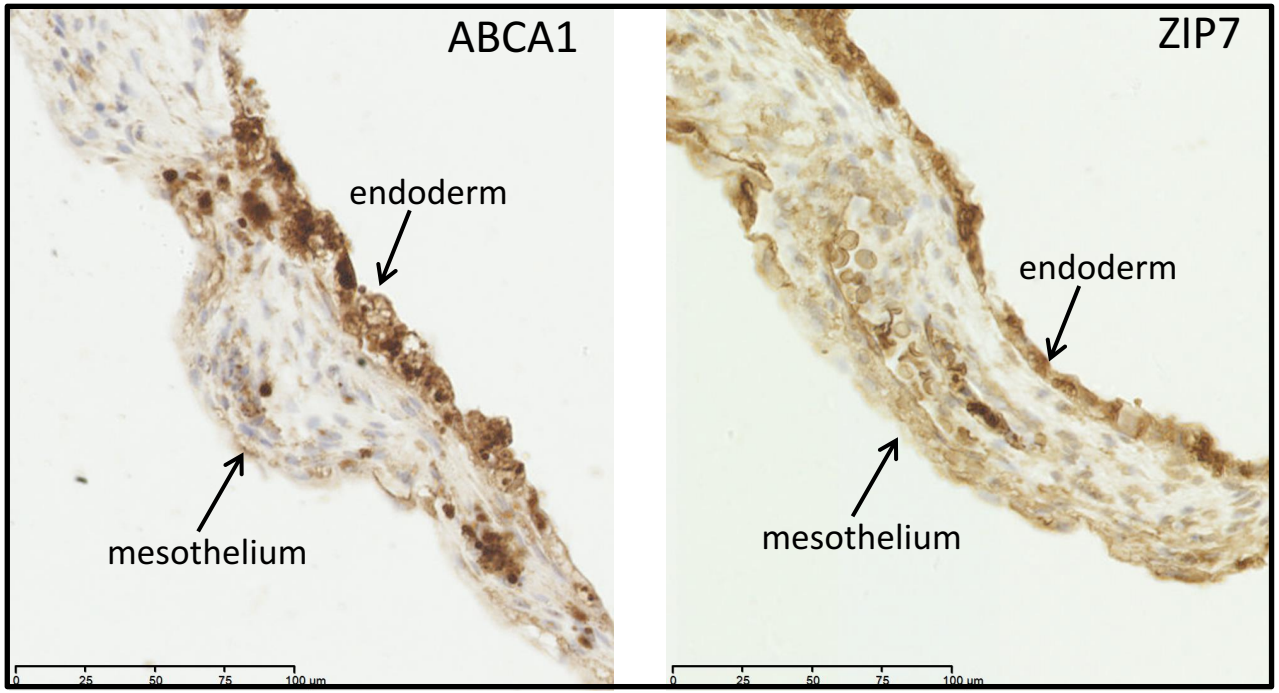
Figure 5 Chord plot connecting GO Biological Process terms associated with "lipid metabolism" and genes encoding transcripts that are shared by the human yolk sac and adult liver. The overlap suggests the yolk sac may perform hepatic functions while the fetal liver differentiates.

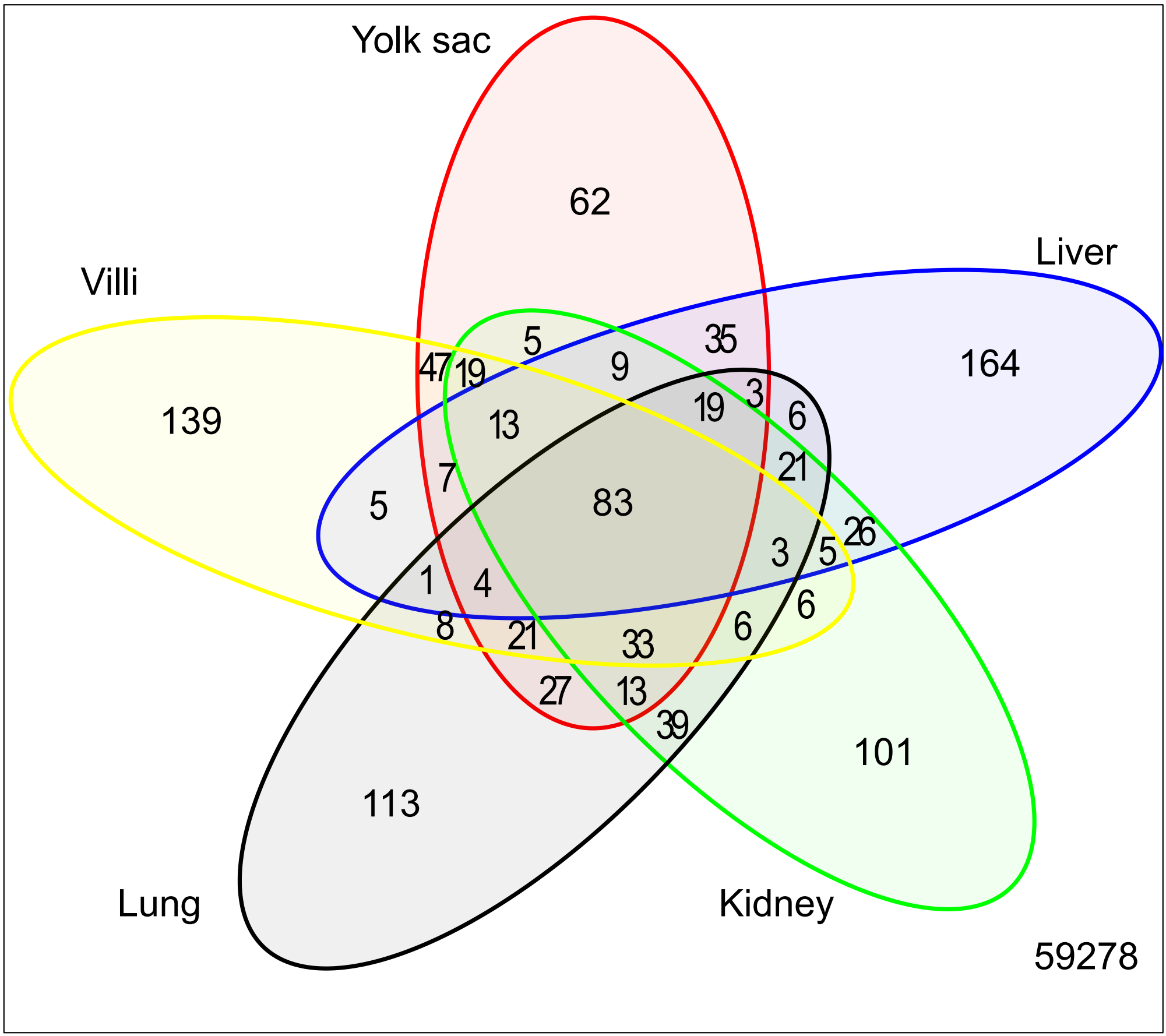
Figure 6 Venn diagrams illustrating the overlap among overrepresented GO terms associated with the 400 most abundant transcripts in each of the human, mouse and chicken yolk sacs: A, "biological process"; B, "cellular component" and C, "molecular function". The considerable overlap among the species in all three categories suggests conservation of functions.

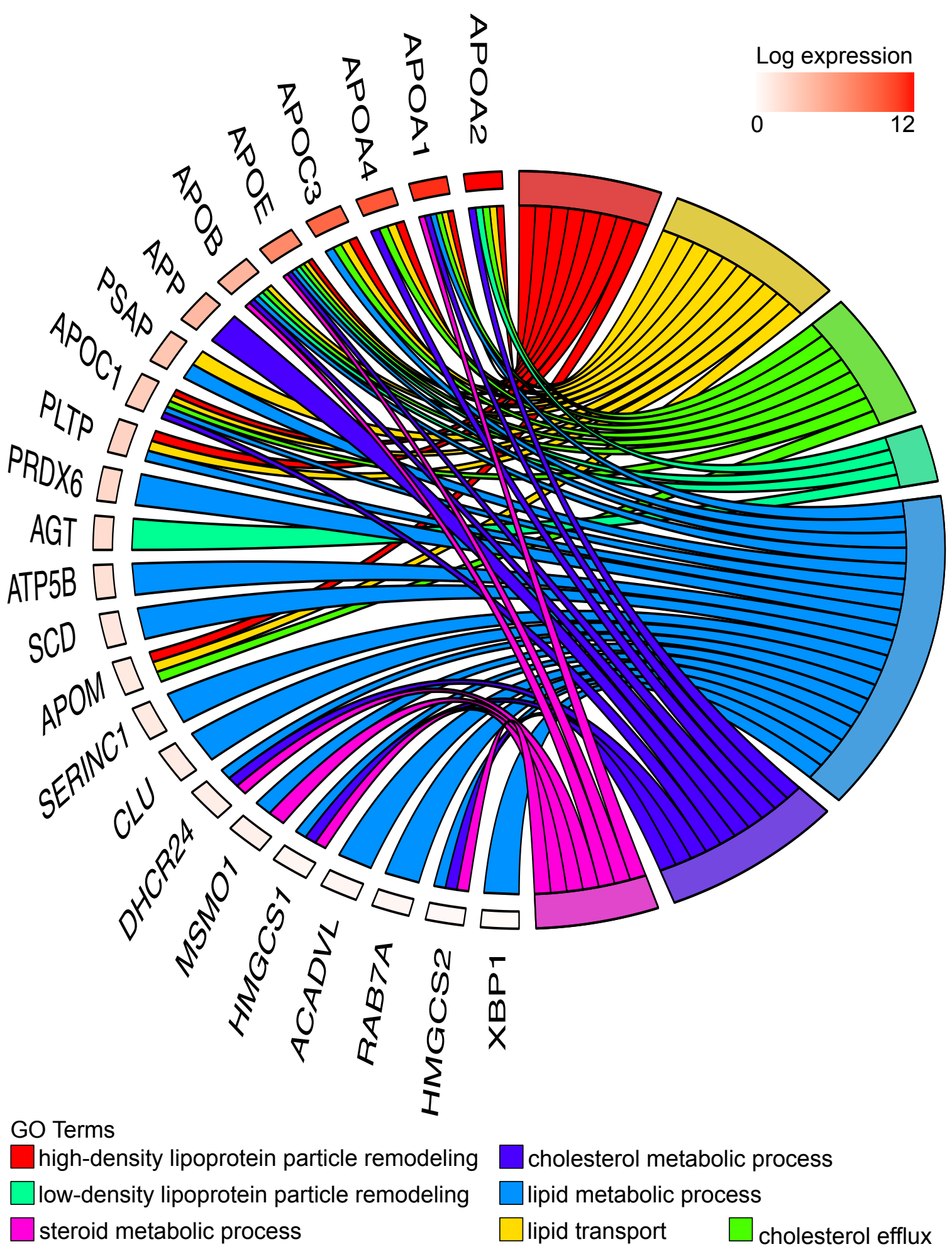
Figure 7 Diagrammatic comparison of the nutrient pathway during early pregnancy in the mouse (A), and the speculated pathway in the human (B). In the mouse, histotrophic secretions (green) released from the endometrial glands (EG) are phagocytosed (1) by the endodermal cells (E) of the visceral layer of the inverted yolk sac (YS). Following fusion with lysosomes (2), digestion of maternal proteins leads to release of amino acids that are transported (3) to the fetal circulation (FC). In the human, histotrophic secretions are released from the endometrial glands through the developing basal plate of the placenta into the intervillous space (IVS), and are phagocytosed (1) by the syncytiotrophoblast (STB) (ref. 42). We speculate that following digestion by lysosomal enzymes (2), free amino acids are transported (3) by efflux transporters to the coelomic fluid (CF) where they accumulate. Nutrients in the CF may be taken up by the mesothelial cells (M) of the yolk sac and transported (4) into the fetal circulation (FC). Alternatively, they may diffuse into the cavity of the yolk sac and be taken up by the endodermal cells (5). Some intact maternal proteins may also be released into the CF by exocytosis of residual bodies (6), and be engulfed by the mesothelial cells (7). CTB; cytotrophoblast cells.

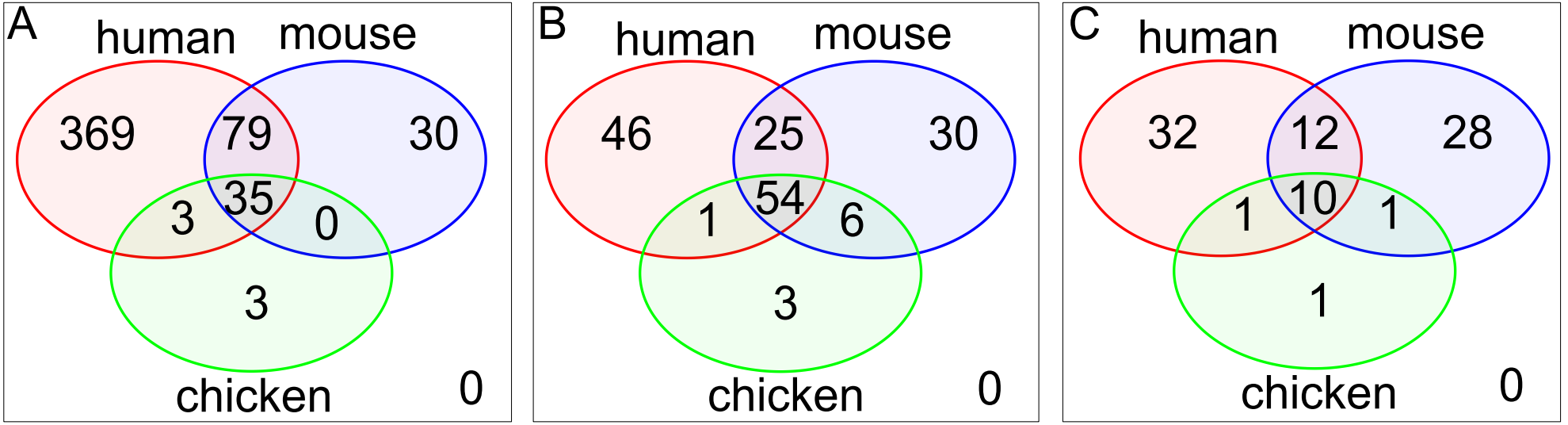

log
expression
6
12
APOA2
APOA1
APOA4
APOC3
APOE
APOB
APP
APOC1
AGT
APOM
CLU
DHCR24
HMGCS1
HMGCS2
XBP1
GO Terms
positive regulation of cholesterol esterification
reverse cholesterol transport
cholesterol metabolic process
cholesterol efflux
cholesterol transport
cholesterol homeostasis

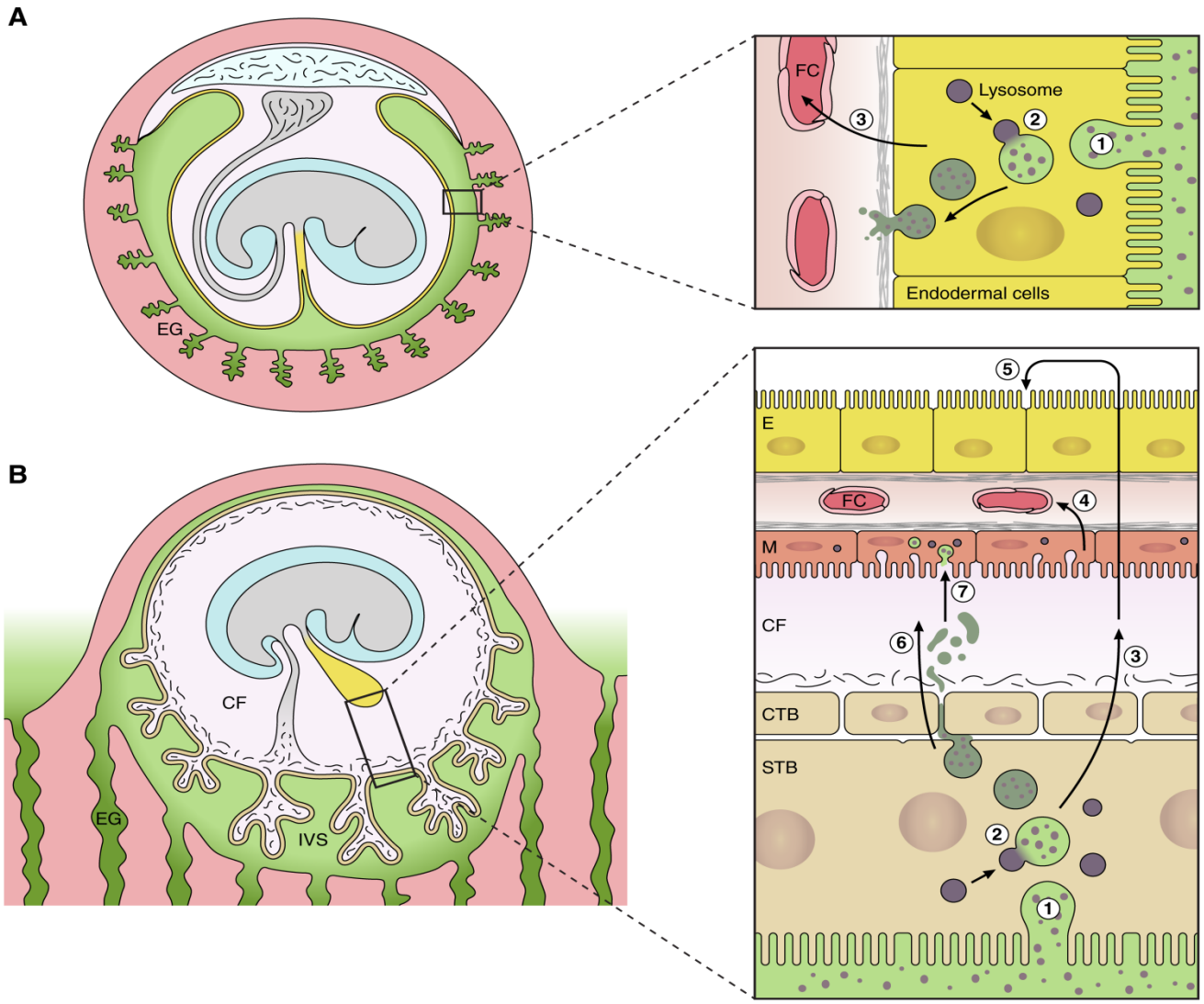

Log expression
3
12
APOA1
APOA4
APOB
AGT
APOE
CLU
PLTP
APOA2
APOD
GM2A
APOM
APOC3
GO Terms
positive regulation of cholesterol esterification
high-density lipoprotein particle assembly
high-density lipoprotein particle clearance
high-density lipoprotein particle remodeling
reverse cholesterol transport
cholesterol transport
lipid transport


ABCA1
endoderm
mesothelium
0
25
50
75
100 um
ZIP7
endoderm
mesothelium
0
25
50
75
100 um


Yolk sac
Villi
Liver
Lung
Kidney
62
139
164
113
101
59278
5
35
47
19
9
19
3
6
13
21
7
5
83
26
3
5
1
4
6
8
21
33
6
27
13
39


Log expression
0
12
APOA2
APOA1
APOA4
APOC3
APOE
APOB
APP
PSAP
APOC1
PLTP
PRDX6
AGT
ATP5B
SCD
APOM
SERINC1
CLU
DHCR24
MSMO1
HMGCS1
ACADVL
RAB7A
HMGCS2
XBP1
GO Terms
high-density lipoprotein particle remodeling
low-density lipoprotein particle remodeling
steroid metabolic process
cholesterol metabolic process
lipid metabolic process
lipid transport
cholesterol efflux


A
human
mouse
369
79
30
35
3
0
3
chicken
0
B
human
mouse
46
25
30
54
1
6
3
chicken
0
C
human
mouse
32
12
28
10
1
1
1
chicken
0


A
EG
FC
Lysosome
3
2
1
Endodermal cells
B
CF
EG
IVS
5
E
FC
4
M
7
CF
6
3
CTB
STB
2
1

Supporting information

RNA-Seq reveals conservation of function among the yolk sacs of human, mouse and chicken

Tereza Cindrova-Davies, Eric Jauniaux, Michael Elliot, Sungsam Gong, Graham J Burton, D. Stephen Charnock-Jones

Methods

Proteomics. 1D gel lanes were cut into 8 bands and each band was transferred into a 96-well PCR plate. The bands were cut into 1mm^2 pieces, destained, reduced (DTT) and alkylated (iodoacetamide) and subjected to enzymatic digestion with trypsin overnight at 37°C. After digestion, the supernatant was pipetted into a sample vial and loaded onto an autosampler for automated LC-MS/MS analysis.

All LC-MS/MS experiments were performed using a Dionex Ultimate 3000 RSLC nanoUPLC (Thermo Fisher Scientific Inc, Waltham, MA, USA) system and a QExactive Orbitrap mass spectrometer (Thermo Fisher Scientific Inc, Waltham, MA, USA). Separation of peptides was performed by reverse-phase chromatography at a flow rate of 300 nL/min and a Thermo Scientific reverse-phase nano Easy-spray column (Thermo Scientific PepMap C18, 2µm particle size, 100Å pore size, 75µm i.d. x 50cm length). Peptides were loaded onto a pre-column (Thermo Scientific PepMap 100 C18, 5µm particle size, 100Å pore size, 300µm i.d. x 5mm length) from the Ultimate 3000 autosampler with 0.1% formic acid for 3 minutes at a flow rate of 10 µL/min. After this period, the column valve was switched to allow elution of peptides from the pre-column onto the analytical column. Solvent A was water + 0.1% formic acid and solvent B was 80% acetonitrile, 20% water + 0.1% formic acid. The linear gradient employed was 2-40% B in 30 minutes.

The LC eluant was sprayed into the mass spectrometer by means of an Easy-Spray source (Thermo Fisher Scientific Inc.). All *m/z* values of eluting ions were measured in an Orbitrap mass analyzer, set at a resolution of 70000 and was scanned between *m/z* 380-1500. Data dependent scans (Top 20) were employed to automatically isolate and generate fragment ions by higher energy collisional dissociation (HCD, NCE:25%) in the HCD collision cell and measurement of the resulting fragment ions was performed in the Orbitrap analyser, set at a resolution of 17500. Singly charged ions and ions with unassigned charge states were excluded from being selected for MS/MS and a dynamic exclusion window of 20 seconds was employed.

Post-run, the data was processed using Protein Discoverer (version 1.4., ThermoFisher). Briefly, all MS/MS data were converted to mgf files and the files were then submitted to the Mascot search algorithm (Matrix Science, London UK) and searched against the UniProt human database (153168 sequences; 54677058 residues). Variable modifications of oxidation (M) and deamidation (NQ) and a fixed modification of carbamidomethyl (C) were

applied and the peptide and fragment mass tolerances were set to 5ppm and 0.1 Da, respectively. A significance threshold value of $p<0.05$ and a peptide cut-off score of 20 were also applied. All the Mascot data from each lane was merged to give a single output file. Finally, Mascot .dat files were entered into the Scaffold software (Proteome Software, Oregon USA) so that differences between the two protein samples could be compared.

Two samples had very few recognizable peptide hits these samples were excluded from the subsequent analysis. Unique hits to identifiable proteins were counted and those with the same gene name in the identifier filed combined. The corresponding gene names for the proteins detected in 4 out of the 5 samples were used in Gene Ontology analysis as described below (Panther).

Bioinformatic analysis

RNA-Seq data processing

Quality assessment and trimming of the de-multiplex reads was carried out using FastQC and cutadapt respectively (http://www.bioinformatics.babraham.ac.uk/projects/fastqc(1). The trimmed short reads were mapped to the human reference genome (hg19) using TopHat2 (version 2.0.12), a splice-aware mapper(2). Uniquely mapped reads were counted using HTSeq (version 0.6.0)(3) and the relative transcript abundance determined using DESeq2 (version 1.6.3)(4).

Other datasets

To compare the levels of transcripts in the yolk sac and placental villi with those in the liver, lung and kidney processed RNA-Seq data was down loaded from the EBI Expression Atlas (version: 0.1.4-SNAPSHOT, experiment E-MTAB-513 http://www.ebi.ac.uk/gxa/experiments/E-MTAB-513?ref=aebrowse).

To compare with 27 other tissues processed RNA-Seq data was downloaded from Expression Atlas - E-MTAB-1733 (http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1733)(5).

GO analysis

The most abundant 400 transcripts in the tissues of interest were identified (ranked by mean RPKM) and gene ontology analyzed by using Panther(6) (http://pantherdb.org release 20160715). The complete Panther data base was used to include the mitochondrially encoded transcripts. The Biological Process, Molecular Function and Cellular Component were all analyzed. The reference sets were all Homo sapiens or Mus musculus genes (known in Panther) and the Bonferroni correction for multiple testing was applied in all cases.
Chord plots of enriched GO terms (Figure 1 for example) were generated in R using GOplot (v 1.0.2. available from CRAN) (7).

Venn diagrams

Venn diagrams were generated in R (vennCounts and VennDiagram, within the Bioconductor limma package) and the p-values for the observed overlaps calculated with SuperExactTest by Wang et al 2015 (v 0.99.2, available from CRAN)(8).

Transcription factor binding sites

Within the most abundant 400 transcripts in the human yolk sac, 19 genes are annotated as "regulation of transcription, DNA-templated" (GO:0006355) including several transcription factors (ATF4, FOS, FOSB, JUN, JUNB and JUND XBP-1 and BHLHE40). In the mouse 8 genes were similarly annotated (Table S3.)

The correlation (Spearman) between each of the transcripts encoding these transcription factors and all other yolk sac transcripts (rpkm >15) was calculated. The transcripts with rho >0.8 were identified. Over-represented transcription factor binding motifs in these genes were identified (JASPAR2016 and TFBSTools). Candidate motifs recognized by the transcription factors (or family for JUN and FOS) were enriched in the 1kb and 5kb upstream of the TSS of the highly correlated genes. Regions of interest were extracted using Biostrings and transcription factor binding motifs identified using JASPAR2016 and TFBSTools, all in R. P values were corrected for multiple testing(9) (Datasets S7 & 8).

Similar analysis was carried out for the mouse data in which 8 genes were annotated as above. Atf4 was the only transcription factor with a binding motif in the JASPAR2016 database. Candidate motifs recognized by Atf4 were enriched in the 1kb and 5kb upstream of the TSS of the highly correlated genes (Datasets S9 &10).

Other manipulations were carried out in R (BiomaRt, dplyr and made4) and Excel.

Results

Sample quality control

Hierarchical clustering of the yolk sac reads indicated that 1 sample did not cluster with any other sample and formed a separate branch. This sample had extremely high levels of villous specific transcripts (CGA, CGB, CGB5, CGB8, PSG1, PSG2, PSG3, PSG4, PSG9, LEP, KISS1 and CSH2; for example CGA ~73,000 vs a mean of ~320 for the other samples). It is therefore likely that the sample is contaminated with villous material. This sample was removed and the level of these transcripts examined in the remaining samples. Two additional samples had markedly different levels of these transcripts and so were also excluded from subsequent analysis.

RNASeq data from the first trimester villus samples were similarly processed and 1 sample clustered alone as a well separated branch. This sample had higher levels of decidual transcripts (such as IGF2, IGFPB4, IGFPB5, IGFPB7 and low levels of villous transcripts (CGA, CGB, CGB5, LEP and KISS1). This sample was therefore excluded from subsequent analysis.

A similar approach was adopted with the mouse yolk sac samples and 1 sample clustered as a well-separated branch. This sample had much lower levels of the fetal hemoglobin transcripts (Hbb-y, Hbb-bh1 and Hba-x) whereas these were the most abundant in all the other samples. This sample was therefore excluded from subsequent analysis.

Mitochondrial transcripts

Various enriched GO terms are associated with mitochondrial activity ("mitochondrial respiratory chain complex", and "mitochondrial ATP synthesis coupled electron transport" for example). This is consistent with the previously reported high density of mitochondria in the yolk sac(10). We calculated the proportion of transcripts present that were encoded by the mitochondrial genome (Figure S1). The fractions for the yolk sac and first trimester villi are slightly above the average and these data are consistent with that showing the kidney and heart had the highest mitochondrial fractions(11).

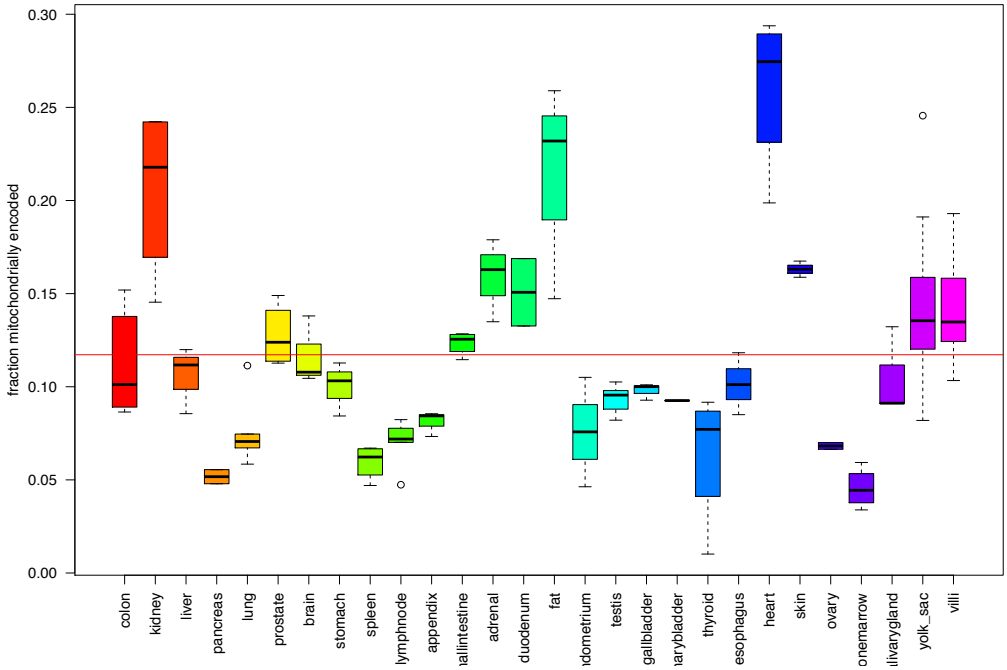


Figure S1. The fraction of mitochondrially encoded transcripts in each tissue type is the proportion mitochondrial genes over the sum over all RPKM values in that tissue

Mitochondrially encoded transcripts directly involved in the electron transfer chain and (or) energy generation were highly correlated with each other (Spearman's rho >0.71) and not with other transcripts in the most abundant 400 (Datasets S6). One other transcript was correlated with these (HNRNPA2B1). Other mitochondrially encoded transcripts (12S and 16S rRNA) were not similarly correlated.

Comparison with other tissues.

We used hierarchical clustering to examine the relationships among the expression patterns of the most variable transcripts detected at rpkm ≥ 0.1 in the yolk sac, placental villi and a range of other tissues. As previously reported, clustering recapitulated the origin of the tissue(5). The first trimester placental villi and the term placental samples clustered together with the yolk sac cluster forming an adjacent branch (Figure S2). This indicates the first and third trimester villi have similar transcript profiles and the 3 extra-embryonic tissues are more similar to each other than to the other tissues examined. The yolk sac samples form 2 closely linked branches, and on closer inspection of the samples falling in these 2 groups, we noted

a marked difference in the level of transcripts encoding ALB, AFP, HBE, HBZ. It is known from previous studies that the size of the yolk sac declines from a maximum diameter of ~6mm at approximately 10 weeks gestation (12). The gestational age of the samples used in this study is only available for a subset of the samples. However, for those where this is known the samples obtained earlier in gestation (7+0, 8+2, 9+0 and 9+2) are on a separate branch to the known later sample (11+0). Male and female samples are distributed evenly across both branches.

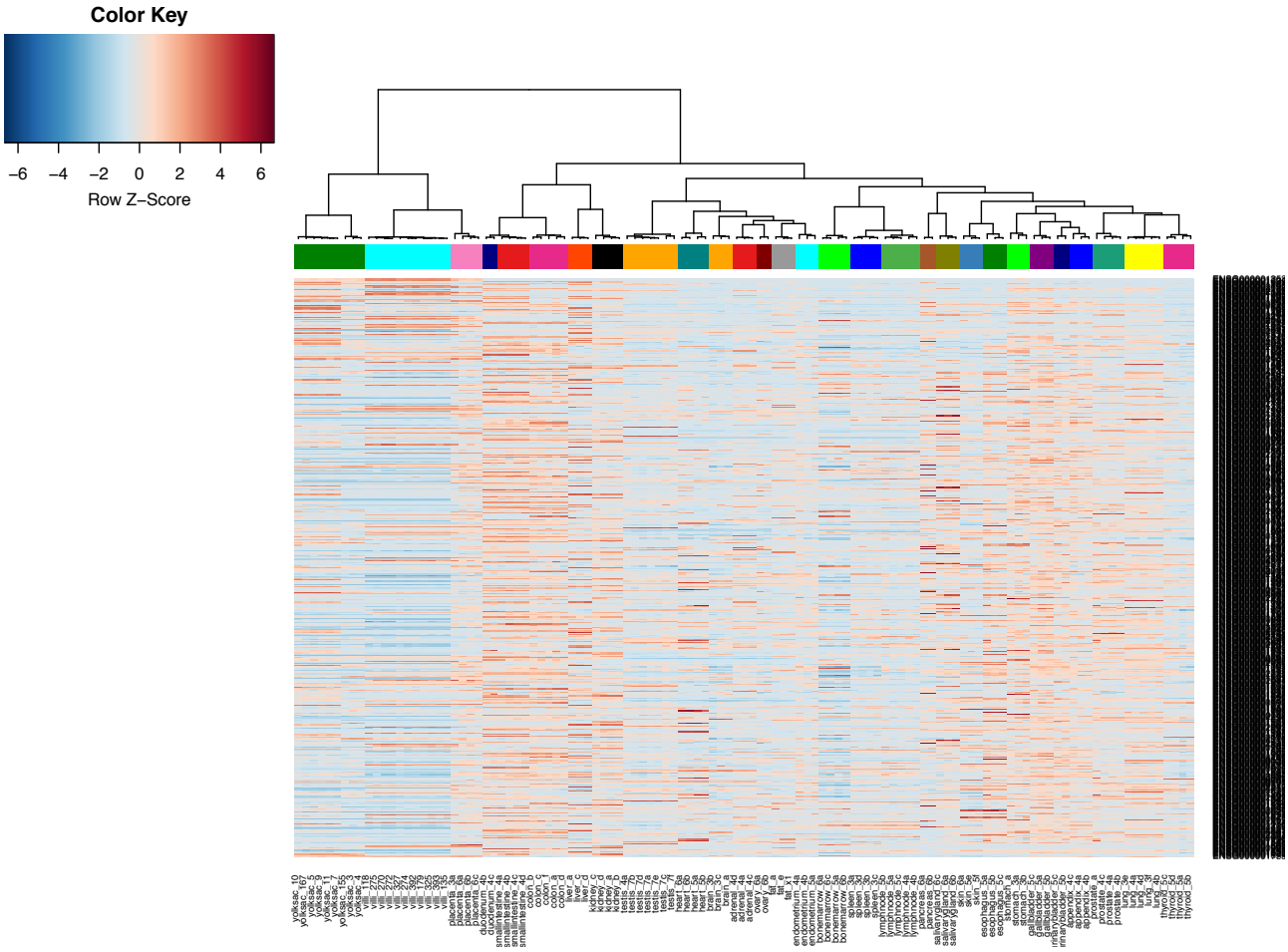
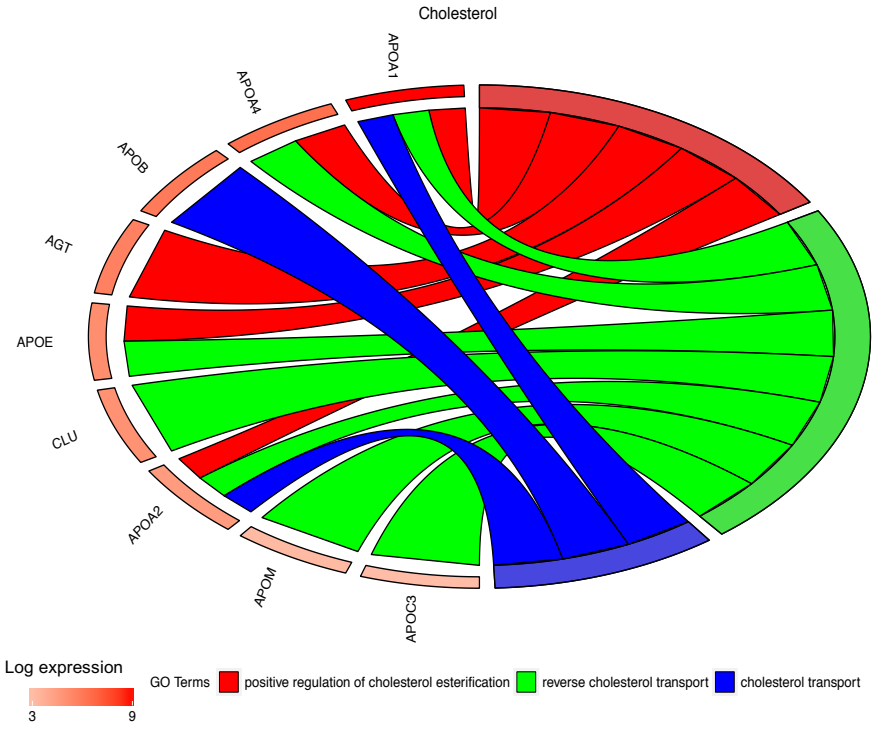


Figure S2. Hierarchical clustering of multiple tissues based on the 500 transcripts with the highest standard deviation(5).



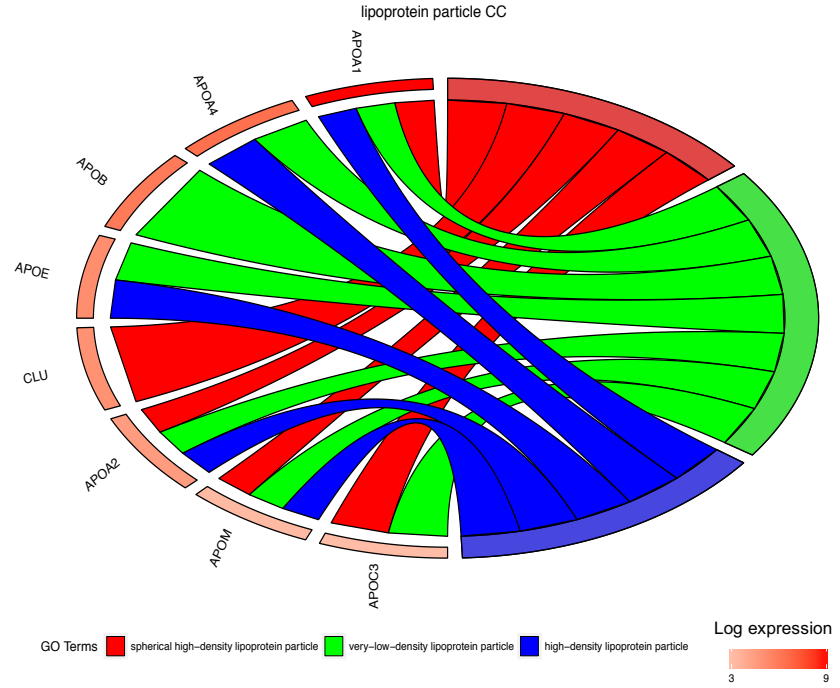


Figure S3 a & b.
Chord plots illustrating the proteins annotated with the GO molecular function and cellular components “cholesterol” and “lipoprotein particle” respectively are present coelomic fluid.

Table S1

Human Yolk sac		Human first trimester villi		Mouse yolk sac	
Sample ID	Reads	Sample ID	Reads	Sample ID	Reads
YS167	48,834,289	P270	4,058,853	B6YS1.2	36,961,551
YS7	63,379,887	P325	23,438,297	B6YS3.1	16,745,203
YS9	58,695,200	P392	12,959,581	B6YS9.1	24,340,906
YS10	65,310,562	P393	58,672,011	B6YS10.1	15,963,038
YS11	37,760,671	P118	76,386,477	B6YS11.1	64,903,212
YS3	39,471,048	P272	205,189,675	B6YS24.1	20,101,598
YS4	28,688,319	P274	30,427,351	B6YS27.1	32,624,089
YS5	33,224,068	P275	131,727,903	B6YS28.1	36,481,077
YS155	39,022,547	P179	22,238,388		
		P327	89,441,895		
median	39,471,048	median	30,427,351	median	28,482,498

Summary of the number of RNASeq reads (single-end 50base) obtained for the human and mouse yolk sacs and the first trimester human placental villi.

Table S2

Common circulating blood proteins	**Serum albumin**, Afamin (albumin family, vit E-binding), Ig (κC, γ-1,-2, λ, V-II, V-III), **Apolipoproteins** (**A-I, A-II, A-IV, B, B100, D, E,** M), clusterin (apolipoprotien J), **heparan sulfate proteoglycan (basement-membrane-specific)**, α-2-macroglobulin, Hemopexin (β-1B-glycoprotein), α-1B-glycoprotein, **α-2-HS-glycoprotein**, **α-1-acid glycoprotein**, **Gelsolin**, **Fibrinogen α**, Vitronectin, **Zinc-α-2-glycoprotein**, Histidine-rich glycoprotein, **β-2-microglobulin, α-1-acid glycoprotein, Leucine-rich α-2-glycoprotein,** β-2-glycoprotein 1, Serum amyloid A-4 protein
Coagulation and complement factors	**Complement** (**B, C3, C4, C7, C9,** D**, H, I**), Kininogen, **Coagulation factors (V, XII)**
Blood transport and binding proteins	Serotransferrin, Ceruloplasmin, **Transthyretin, Vitamin D-binding protein, Retinol-binding protein**, Galectin-3-binding protein, Latent-transforming growth factor beta-binding protein, Phospholipid transfer protein, **Hemoglobin (α, β**, ε, ζ), **Thyroxine-binding globulin**, Insulin-like growth factor-binding protein 3, **Latent-transforming growth factor β-binding protein, Corticosteroid-binding globulin**
Protease inhibitors	**α-1-antitrypsin**, Antithrombin-III, **α-1-antichymotrypsin**, **Inter-α-trypsin inhibitor**, **Plasma protease C1 inhibitor**, **Inter-α-trypsin inhibitor**, α-2-antiplasmin, Cystatin c (potent inhibitor of lysosomal proteinases), **Heparin cofactor 2** (rapidly inhibits thrombin in the presence of dermatan sulfate or heparin), inhibits trypsin, plasmin, and lysosomal granulocytic elastase), Metalloproteinase inhibitor 1

Proteases and other enzymes	**Plasminogen, angiotensinogen**, Procollagen C-endopeptidase enhancer, Glutathione peroxidase, Prothrombin, Cu-Zn superoxide dismutase,
Cytokines and hormones	α-fetoprotein, Pigment epithelium-derived factor (serpin 1), Choriogonadotropin subunit β
Channel and receptor-derived peptides	
Miscellaneous (structural, nuclear etc.)	Fibronectin, Fibulin-1, Decorin (proteoglycan), **Collagen**s-α1, -2,-3, Laminin, Versican core protein, Periostin, **Keratins (I, II)**, Transforming growth factor-β-induced protein, Lumican (proteoglycan), Laminin (β1, γ1), glycodelin, Hyaluronan and proteoglycan link protein, nidogen-1 (entactin, basement memb), titin (striated muscle contraction), actin, Fibulin-2, Osteonectin (SPARC, a glycoprotein in the bone that binds calcium), Spondin-1 (axon guidance), Extracellular matrix protein 1, Dystroglycan (dystrophin-associated glycoprotein), Cadherin-1 (transmembrane protein)

Proteins in **bold** are also reported in human serum

Categorization of the proteins detected in the human coelomic fluid and comparison with serum proteins

Table S3

Human

ensembl gene id	hgnc symbol	description	GO linkage type
ENSG00000065978	YBX1	Y box binding protein 1	IDA
ENSG00000128272	ATF4	activating transcription factor 4	IEA, ISS
ENSG00000145741	BTF3	basic transcription factor 3	IEA
ENSG00000009307	CSDE1	cold shock domain containing E1, RNA-binding	IEA
ENSG00000156508	EEF1A1	eukaryotic translation elongation factor 1 alpha 1	IEA
ENSG00000170345	FOS	FBJ murine osteosarcoma viral oncogene homolog	IEA
ENSG00000152795	HNRNPDL	heterogeneous nuclear ribonucleoprotein D-like	IEA
ENSG00000171223	JUNB	jun B proto-oncogene	IEA
ENSG00000130522	JUND	jun D proto-oncogene	IEA
ENSG00000177606	JUN	jun proto-oncogene	IEA
ENSG00000107438	PDLIM1	PDZ and LIM domain 1	IEA
ENSG00000177469	PTRF	polymerase I and transcript release factor	IEA
ENSG00000149273	RPS3	ribosomal protein S3	IEA
ENSG00000179218	CALR	calreticulin	TAS
ENSG00000169714	CNBP	CCHC-type zinc finger, nucleic acid binding protein	TAS
ENSG00000107223	EDF1	endothelial differentiation-related factor 1	TAS
ENSG00000089009	RPL6	ribosomal protein L6	TAS

ENSG00000134107	BHLHE40	basic helix-loop-helix family, member e40	NAS
ENSG00000167244	IGF2	insulin-like growth factor 2 (somatomedin A)	NAS

Mouse

ensembl gene id	mgi symbol	description	GO linkage type
ENSMUSG00000042406	Atf4	activating transcription factor 4	IDA, IMP
ENSMUSG00000006932	Ctnnb1	catenin (cadherin associated protein), beta 1	IDA
ENSMUSG00000020267	Hint1	histidine triad nucleotide binding protein 1	IEA, ISS, ISO
ENSMUSG00000028639	Ybx1	Y box protein 1	IEA, ISS, ISO
ENSMUSG00000021660	Btf3	basic transcription factor 3	IEA
ENSMUSG00000051223	Bzw1	basic leucine zipper and W2 domains 1	IEA
ENSMUSG00000050966	Lin28a	lin-28 homolog A (C. elegans)	IEA
ENSMUSG00000055839	Tceb2	transcription elongation factor B (SIII), polypeptide 2	IEA

Table of the genes annotated with the GO term “regulation of transcription, DNA-templated" (GO:0006355) in the most abundant human and mouse yolk sac transcripts. GO evidence codes (go_linkage_type) are also provided (see http://geneontology.org/page/guide-go-evidence-codes)

Table S4

BP	BP description	MF	MF description	CC	CC description
GO:0006364	rRNA processing	GO:0003676	nucleic acid binding	GO:0000323	lytic vacuole
GO:0006412	translation	GO:0003723	RNA binding	GO:0005575	cellular_component
GO:0006518	peptide metabolic process	GO:0003735	structural constituent of ribosome	GO:0005576	extracellular region
GO:0006807	nitrogen compound metabolic process	GO:0005198	structural molecule activity	GO:0005615	extracellular space
GO:0008152	metabolic process	GO:0005488	binding	GO:0005622	intracellular
GO:0009058	biosynthetic process	GO:0016491	oxidoreductase activity	GO:0005623	cell
GO:0009059	macromolecule biosynthetic process	GO:0019843	rRNA binding	GO:0005730	nucleolus
GO:0009123	nucleoside monophosphate metabolic process	GO:0044822	poly(A) RNA binding	GO:0005737	cytoplasm
GO:0009161	ribonucleoside monophosphate metabolic process	GO:0097159	organic cyclic compound binding	GO:0005739	mitochondrion
GO:0009987	cellular process	GO:1901363	heterocyclic compound binding	GO:0005740	mitochondrial envelope
GO:0010467	gene expression			GO:0005746	mitochondrial respiratory chain
GO:0016072	rRNA metabolic process			GO:0005764	lysosome
GO:0019538	protein metabolic process			GO:0005773	vacuole
GO:0022613	ribonucleoprotein complex biogenesis			GO:0005829	cytosol
GO:0034641	cellular nitrogen compound metabolic process			GO:0005840	ribosome
GO:0034645	cellular macromolecule biosynthetic process			GO:0005912	adherens junction

GO:0042254	ribosome biogenesis
GO:0042255	ribosome assembly
GO:0043043	peptide biosynthetic process
GO:0043603	cellular amide metabolic process
GO:0043604	amide biosynthetic process
GO:0044085	cellular component biogenesis
GO:0044237	cellular metabolic process
GO:0044238	primary metabolic process
GO:0044249	cellular biosynthetic process
GO:0044267	cellular protein metabolic process
GO:0044271	cellular nitrogen compound biosynthetic process
GO:0044281	small molecule metabolic process
GO:0044710	single-organism metabolic process
GO:0055114	oxidation-reduction process
GO:0065003	macromolecular complex assembly
GO:0071704	organic substance metabolic process
GO:1901564	organonitrogen compound metabolic process
GO:0005924	cell-substrate adherens junction
GO:0005925	focal adhesion
GO:0015934	large ribosomal subunit
GO:0015935	small ribosomal subunit
GO:0016020	membrane
GO:0022625	cytosolic large ribosomal subunit
GO:0022626	cytosolic ribosome
GO:0022627	cytosolic small ribosomal subunit
GO:0030054	cell junction
GO:0030055	cell-substrate junction
GO:0030529	intracellular ribonucleoprotein complex
GO:0031090	organelle membrane
GO:0031966	mitochondrial membrane
GO:0031982	vesicle
GO:0031988	membrane-bounded vesicle
GO:0032991	macromolecular complex
GO:0043226	organelle

	organonitrogen compound biosynthetic		
GO:1901566	process	GO:0043227	membrane-bounded organelle
	organic substance		
GO:1901576	biosynthetic process	GO:0043228	non-membrane-bounded organelle
		GO:0043229	intracellular organelle
		GO:0043230	extracellular organelle
		GO:0043231	intracellular membrane-bounded organelle
		GO:0043232	intracellular non-membrane-bounded organelle
		GO:0044391	ribosomal subunit
		GO:0044421	extracellular region part
		GO:0044422	organelle part
		GO:0044424	intracellular part
		GO:0044444	cytoplasmic part
		GO:0044445	cytosolic part
		GO:0044446	intracellular organelle part
		GO:0044455	mitochondrial membrane part
		GO:0044464	cell part
		GO:0070062	extracellular exosome
		GO:0070161	anchoring junction
		GO:0070469	respiratory chain
		GO:0098798	mitochondrial protein complex
		GO:0098800	inner mitochondrial membrane protein complex
		GO:1903561	extracellular vesicle

Over represented GO terms shared among the 400 most abundant transcripts in the human, mouse and chicken yolk sacs. BP, Biological process; MF molecular function and CC, Cellular component

References

1. Martin M (2016) Cutadapt removes adapter sequences from high-throughput sequencing reads. 1–3 doi: 10.14806/ej.17.1.200.

2. Kim D, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14(4):R36.

3. Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.

4. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.

5. Fagerberg L, et al. (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* 13(2):397–406.

6. Mi H, Muruganujan A, Thomas PD (2013) PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* 41(Database issue):D377–86.

7. Walter W, Sánchez-Cabo F, Ricote M (2015) GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics* 31(17):2912–2914.

8. Wang M, Zhao Y, Zhang B (2015) Efficient Test and Visualization of Multi-Set Intersections. *Sci Rep* 5:16923.

9. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B* 57(1):289–300.

10. Pereda J, et al. (1994) The structure of the human yolk sac: a scanning and transmission electron microscopic analysis. *Arch Histol Cytol* 57(2):107–117.

11. Melé M, et al. (2015) Human genomics. The human transcriptome across tissues and individuals. *Science* 348(6235):660–665.

12. Jauniaux E, Jurkovic D, Henriet Y, Rodesch F, Hustin J (1991) Development of the secondary human yolk sac: correlation of sonographic and anatomical features. *Hum Reprod* 6(8):1160–1166.