Novel and divergent genes in the evolution of placental mammals

Thomas L. Dunwell[1], Jordi Paps[1,2] and Peter W.H. Holland[1*]

1 – University of Oxford, Department of Zoology.

2 - University of Essex, School of Biological Sciences.

*** Corresponding author: peter.holland@zoo.ox.ac.uk**

Key words: New genes; molecular evolution; MCL clustering; Eutheria; Placentalia

Running head: Genes of placental mammals

Abstract

Analysis of genome sequences within a phylogenetic context can give insight into the mode and tempo of gene and protein evolution, including inference of gene ages. This can reveal whether new genes arose on particular evolutionary lineages and were recruited for new functional roles. Here, we apply MCL clustering with all-vs-all reciprocal BLASTP to identify and phylogenetically date 'Homology Groups' amongst vertebrate proteins. Homology Groups include new genes and highly divergent duplicate genes. Focussing on the origin of the placental mammals within the Eutheria, we identify 357 novel Homology Groups that arose on the stem lineage of Placentalia, 87 of which are deduced to play core roles in mammalian biology as judged by extensive retention in evolution. We find the human homologues of novel eutherian genes are enriched for expression in preimplantation embryo, brain, and testes, and enriched for functions in keratinization, reproductive development, and the immune system.

Introduction

Living mammals are divided into three major clades: monotremes, marsupials, and placentals. The placental mammals are the most speciose of the three with almost 4000 described species encompassing a striking range of morphological diversity from bats to whales, and elephants to humans.

The common ancestor of placentals and marsupials dates to ~140 to 191 million years ago (mya), whereas the crown Placentalia dates to only ~72 to 107 mya with the oldest fossil at 65 mya [1,2]. Despite the uncertainties (and controversies), these dates suggest a long period of 60 to 80 million years during which the genetic changes occurred that distinguish living placental mammals from marsupials or monotremes.

The inclusive clade (total group) encompassing crown Placentalia and their closest extinct relatives is termed Eutheria and its members can be distinguished from the Metatheria, including marsupials, by several skeletal and dentition characters. Additional physiological and reproductive features are evident in living placental mammals including extended gestation, a well-developed placenta and loss of epipubic bones enabling abdominal expansion during pregnancy. In association with these changes, development of an invasive placenta posed new immunological challenges for placental mammals [3], while reorganisation of blastocyst development is associated with early specification of trophoblast cells [4,5]. Hence, over the interval from the origin of the Eutheria to the origin of the placental mammals a suite of phenotypic characters arose which were exploited by evolution as the Placentalia radiated extensively and colonized a vast range of habitats.

We aim to understand the origin of placental mammals at the molecular level. Genomic changes that could contribute to phenotypic change include changes to cis-regulatory DNA, changes to repetitive DNA landscapes, and the origin and loss of coding and non-coding genes. In addition, co-option of genes from integrated retroviruses has been shown to be important in eutherian mammal evolution, generated *syncytin* genes deployed to facilitate cellular fusion during placentation [6]. Here we investigate the extent to which novel protein-coding genes arose on the stem lineage of the placental mammals, during the first ~60-80

million years of eutherian evolution, and whether novel genes likely contributed to the emergence of the unique phenotypic characters of placental mammals. We define novel genes as including gene duplicates that have undergone unusually extensive sequence change compared to the other gene duplicate (referred to as asymmetric evolution [7]) as well as new genes generated by more complex genomic events (transposition, inversion and repurposing of non-coding DNA).

We describe a comparative analysis of all protein-coding genes present in the genomes of a phylogenetically diverse set of ten eutherian mammals, three non-eutherian mammals (marsupials and monotremes), four reptiles/birds, one amphibian, and two actinopterygian species. Using a recently developed pipeline combining reciprocal all-vs-all BLASTP and Markov Cluster (MCL) grouping on the basis of sequence similarity, we group protein-coding genes into 'Homology Groups' dated to phylogenetic nodes. We identify 357 novel Homology Groups arising on the stem lineage of Placentalia, a subset of 87 of these are extensively maintained across subsequent evolution. Expression profiles and functional annotation suggest recruitment of novel genes to preimplantation embryo, brain, testis, keratinization, and immune functions.

Material and Methods

Protein Data sets

Twenty vertebrate species were chosen on the basis of completeness of genome sequencing and annotation, covering the phylogenetic diversity of placental mammals and a series of nested outgroups. A non-redundant protein dataset for each species was generated by combining NCBI RefSeq and Ensmbl predictions as follows. RefSeq protein data were downloaded from NCBI (accessed July 2015) and filtered to retain only the longest canonical peptide associated with each Entrez gene ID. Protein predictions were also obtained from Ensembl (except *Chrysemys picta* with no Ensembl annotation) and redundancy with RefSeq reduced by removing proteins with matching Entrez gene ID, BLASTP hits of p-value = 0 or 100% BLASTP matches across alignable regions, to generate a final combined proteome for each species (Figure 1). The total data set comprised 468,298 peptide sequences (Electronic Supplementary Material Tables S1, S2).

BLAST-MCL pipeline and data filtering

A local database was created using the combined NCBI-Ensembl protein datasets and reciprocal all-vs-all protein BLASTP searches performed with default settings and a cut off p-value of $5e^{-5}$ using BLASTP version 2.2.27 [8] The output was passed to mcxdeblast with the options '--m9' and '--line-mode=abc' to generate an MCL-compatible format. MCL (version 12-135 [9]) was then used to infer groups of putative homology using the following options '--abc -I 2'; this generates clouds of closely related proteins with significant difference from neighbouring clouds (Figure 2). A Homology Group (HG) was inferred to represent a Novel Ancestral Placental gene, meaning it was present in the last common ancestor of crown Placentalia, if proteins within the cluster were present within one or more Atlantogenata species (Xenarthra or Afrotheria) and one or more Boreoeutheria (Euarchontoglires or Laurasiatheria), and in no outgroups. A subset of these, termed Novel Core Placental genes, were defined as HG present in all (or all but one) of the placental mammal species (Figure 2). Using proteins from the Novel Core Placental clusters, web-based BLASTP searches against the NCBI non-redundant protein sequence database were used to test for false positives resultant from incomplete taxon sampling. Custom scripts used for data filtering are available through GitHub [10].

Phylogenetics

Phylogenetic analysis of all proteins within Novel Core Placental HG used alignments generated with MAFFT (with '--inputorder --anysymbol --ep 0 --maxiterate 1000 --retree 1 –globalpair' options [11]), trimming with trimAl (with the '-automated1' option [12]) and maximum likelihood analysis using FastTree (with '-wag -gamma' options [13]). For species trees, selected proteins were aligned and trimmed as above, and concatenated. Gaps were retained when a protein was absent from a species. Concatenated sequences were analysed using Phylobayes (options '-cat -gtr -nchain 2 100 0.3 50' [14]) and allowed to generate 200,000 trees; consensus trees were obtained by using readpb with a burn in of 1000 trees and subsequent sampling every 10 trees.

GO Pathway and Functional Enrichment

The online web portal DAVID 6.8 (https://david.ncifcrf.gov/ [15]) was used to assess KEGG pathway and GO term annotation enrichment.

RNASeq, Heatmaps, and Expression Clustering

FPKM expression data were generated using CUFFLINKS [16] with default parameters applied to a previously described human tissue expression panel [17]. FPKM values were normalised against the cell or tissue type in which each individual gene was most highly expressed; FPKM values below 2 were treated as 0. Heatmaps were generated in R using the heatmap.2 function of the gplots package and a normalised expression scale of 0-1. Clusters of highly expressed genes were identified by manual inspection of the generated heatmaps.

Results

Identification of Novel Ancestral Placental genes

To investigate gene origin during mammalian evolution, we focussed on well-annotated genome sequences from 10 placental mammals and 10 species representing five nested outgroup clades: marsupials (gray short-tailed opossum and Tasmanian Devil), monotremes (platypus), sauropsids (2 reptiles and 2 birds), amphibians (*Xenopus*), and actinopterygian fish (Figure 1). The placental species include representatives of the four extant major clades:

Euarchontoglires (human, mouse, rabbit), Laurasiatheria (cow, cat, horse, shrew), Afrotheria (elephant, tenrec) and Xenarthra (armadillo). To obtain maximally representative proteome predictions for each genome, we combined NCBI RefSeq and canonical Ensembl protein predictions generating a total dataset of 468,298 peptides. These were used in reciprocal BLASTP searches to identify sequence similarities and the output analysed using MCL to identify groups of putatively homologous proteins (adapted from ref [18]). Although sequence similarity is evident between some groups, these groupings can be considered distinct genes or sets of genes.

Of the total of 20,363 groups of homologous proteins identified (Homology Groups, HG), 5088 are present only in one or more placental mammals species. Using a phylogenetic tree of placental mammals [19] that places Xenarthra as sister group to Afrotheria (collectively the Atlantogenata), and Euarchontoglires sister group to Laurasiatheria (forming Boreotheria), we infer that 9465 HG were present in the last common ancestor of extant placental mammals (i.e. HG present in at least one member of Atlantogenata and at least one Boreoeutheria. or present in at least one non-placental mammal and one placental mammal; Figure 2). Of these, 357 are specific to the placental mammals, not present in any non-placental mammals or other vertebrates. We term these 357 HG 'Novel Ancestral Placental' genes; we infer these genes arose within Eutheria on the stem lineage of Placentalia. The human genome contains genes belonging to 249 of these 357 HG, totalling 376 different genes. Electronic Supplementary Material Table S1 gives accession numbers for each protein assigned to Novel Ancestral Placental HG; Electronic Supplementary Material Table S3 gives numbers of proteins per HG.

To test if new eutherian-specific genes are enriched or depleted across human chromosomes, we compared the number of Novel Ancestral Placental genes (376 genes from 249 HG) located on each human chromosome to the total number of protein-coding genes used in our data set found on each chromosome (Figure 3). Chromosomes 20, Y and X show significant overrepresentation of Novel Ancestral Placental genes (p-values $2.9e^{-3}$, $6.3e^{-4}$ and $6.6e^{-30}$ respectively; Fishers exact test); Chromosome 2 shows depletion (p-value $4e^{-3}$; Fishers exact test).

Identification of Novel Core Placental genes

Of the 357 Novel Ancestral Placental HG, 87 are present in all, or all but one, of the eutherian mammal species analysed (Figure 2). On the basis of extensive retention in subsequent evolution, we infer that these 87 HG contain eutherian-specific proteins expected to be central for 'making a placental mammal'. We therefore term these 87 HG 'Novel Core Placental' genes (Figure 2). Novel Core Placental HG are a subset of the Novel Ancestral Placental HG. In the human genome, 86 of the 87 core HG are present, containing 133 different proteins. Examining the chromosomal distribution of the human representatives reveals that chromosomes 20 and X also have overrepresentation for Novel Core Placental genes (p-values $3.2e^{-4}$ and $3e^{-29}$ respectively; Fishers exact test).

The number of predicted proteins present in each Novel Core Placental HG can vary over tenfold between species; for example, HG648 (encoding membrane-anchored ligands for immune-associated NKG2D activating receptor) contains 2 proteins in *Felis catus* and 31 in *Bos taurus*. Electronic Supplementary Material Table S1 gives accession numbers for each protein assigned to Novel Core Placental HG; Electronic Supplementary Material Table S3 gives numbers of proteins per HG.

The extensive retention of Novel Core Placental genes enables a test of their inferred homology. If Homology Group assignment is accurate, we expect that a phylogenetic tree constructed from sequence alignment should recover the known evolutionary tree for the ten placental mammals in the dataset. First, we used phylogenetic analysis of each HG individually to determine if any contained multiple genes in the most recent common ancestor of extant placental mammals. For 78 of the 87 Novel Core Placental HG these trees were consistent with descent from a single gene, in 6 cases the trees implied descent from 2 genes (indicating that gene duplication had occurred on the placental stem lineage), 2 HG were derived from 3 genes and 1 HG was derived from 5 genes. If a species had experienced additional gene duplications, the gene with the shortest branch length was used. The 101 representative proteins were then aligned, trimmed, and concatenated to generate an alignment of length 26,018 amino acids (Electronic Supplementary Material Table S4). Bayesian phylogenetic analysis of the concatenated alignment recapitulated the expected phylogenetic relationships for the 10 placental mammals (Figure 4).

Functional inference by annotation

To gain insight into possible functions of the Novel Ancestral Placental and Novel Core Placental HG proteins, Gene Ontology (GO) terms and KEGG pathway enrichment was performed using the human genes from each HG (Figure 5).

Of the 133 human genes belonging to 86 (of 87) Novel Core Placental HG, 116 (87%) were assigned one or more GO terms. Among biological processes, functional category enrichment was found for negative transcriptional regulation, keratinization, and natural killer cell-mediated cytotoxicity. In the molecular function category, there is enrichment for proteins involved in WW domain binding and natural killer cell lectin-like receptor binding.

Of the 376 genes from 249 Novel Ancestral Placental HG, 249 (66%) were assigned one or more terms relating to cellular component, biological process, or molecular function. Enrichment was seen for a similar selection of terms, with the addition of male gonad development, spermatogenesis, innate immunity, and defence response to bacteria. Both Novel Ancestral Placental and Novel Core Placental HG proteins were also enriched for pathway functions related to natural killer cell-mediated cytotoxicity (Figure 5).

Functional inference by gene expression

Specificity of gene expression can give insight into the deployment of genes into specific biological processes roles. We therefore examined tissue specificity of gene expression for 336 human genes belonging to Novel Ancestral Placental HG (including Novel Core Placental HG), using publicly available RNA-Seq data from 59 normal human adult and embryo cell types and tissues. Expression values for each gene were normalised against the tissue or cell type in which each gene is most highly expressed, and data clustered to identify groups of genes with similar expression patterns (Figure 6). Normalising ensures that genes with similar biological profiles are clustered, regardless of absolute expression levels. Raw FPKM and normalised data are available in Electronic Supplementary Material Table S5.

Clustering revealed a series of visually distinct groups of genes sharing similar expression profiles, revealing sets of genes likely involved in a range of possible biological processes (Figure 6; Electronic Supplementary Material Figure S1). Groups vary in size from a single gene (e.g. *APOC4* expressed in liver only) to 61 genes (testis). We identify seven clusters of novel placental genes associated with reproductive tissues and pre-implantation embryos (testes, 61 genes; 8-cell and morula, 31 genes; 8-cell embryo only, 14 genes; oocyte, zygote, 2-cell

and 4-cell, 12 genes; embryonic stem cells, 6 genes; late blastocyst, 4 genes; Fallopian tubes, 4 genes). We also note sets of novel placental genes associated with the immune system (9 genes), breast tissue (5 genes), and brain (41 genes), and a set of genes expressed broadly in the majority of tissues examined (23 genes). The identity of genes in highlighted expression sets are given in Table 1; all gene names are present in Electronic Supplementary Material Figure S1.

Most expression sets include genes from the widely-retained Novel Core Placental HG, as well as other Novel Ancestral Placental HG. Interestingly, the brain expression set is significantly enriched in Novel Core Placental genes (p-value = $4e^{-4}$).

Evolutionary origin of novel genes

Reconstructing the mutational pathways that gave rise to each novel placental gene is complicated by the length of the elapsed time since their origin. To investigate if sequence divergence and/or gene duplication underpinned origin, we examined sequence relationships between HG using reciprocal BLASTP. For the majority of Novel Core Placental HG, we detected no BLASTP hits to any other Novel Core Placental HG (Figure 7A). The exceptions were: (1) five putatively related HG encoding TCEAL and BEX proteins (InterPro IPRO21156); (2) two HG encoding a subset of chromosomally-clustered WFDC proteins; (3) three HG encoding retroposon Gag-like proteins; and (4) two HG encoding KRTAP keratin-associated proteins (ID1-4 in Figure 7A and Electronic Supplementary Material Table S6).

Expanding the BLASTP analysis to all HG was used to search for additional evolutionary relationships (Figure 7B). This revealed that 33 of the Novel Core Placental HG have no significant BLASTP similarity to any HG outside of placental mammals. A total of 15 Novel Core Placental HG have sequence similarity to other HG found across placental and non-placental mammals, and a further 39 have sequence similarities more broadly than mammals the most extreme being HG9135 (ID 5 in Figure 7) with blast hits to 26 other HG (Electronic Supplementary Material Table S7). The degree of sequence similarity to proteins outside placental mammals is far lower than the similarities within the placental HG indicating relationship to a broader protein superfamily. For example, Novel Core Placental HG 3030 has two proteins in human, CYS9 and CYS9L, comprising the Cystatin 9 family of proteases; the cystatin gene superfamily is found across eukaryotes, but the Cystatin 9 family has previously

been shown to be specific to placental mammals [20]. Similarly, Novel Core Placental HG 648 has six proteins in humans comprising the ULBP/RAET family of MHC Class I-related proteins, which are distantly related to genes in marsupials [21].

To further trace origins, we focussed on all Novel Core Placental HG that were single copy in all eutherian mammals, and compared genomic position and organisation in human to the syntenic region in opossum. These comparisons suggested four distinct mutational routes for the origin of Novel Placental HG: (1) extensive sequence divergence of a pre-existing gene; (2) tandem gene duplication followed by asymmetric sequence divergence from a pre-existing gene; (3) origin of a protein-coding gene in a location where no gene is present in non-eutherian mammals; and (4) genomic rearrangement associated with the origin of a putatively novel sequence. Not all genes could be clearly assigned to just one of these categories. Examples of these four routes are given in Figure 8.

Discussion

Although much attention in comparative biology is focussed on genes and genetic pathways that are shared between species, it is also clear that there has been much novelty in evolution. For example, as each new genome sequence is reported, suites of genes are discovered without clear homologues in other species, suggesting a high rate of novelty. It could be argued that our vision of novelty is exaggerated because in many cases genomes are being compared that are distantly related, but the conclusion cannot be escaped that many new genes arise in evolution. Putting numbers or rates onto novelty is difficult, however, since there is no single definition of what constitutes a new gene. At one extreme, focus could be restricted only to genes that emerged by de novo origin from non-coding sequence [22], or alternatively one could include those originated by assembly from disparate domain components or by radical sequence divergence with or without duplication [7]. Mechanistic definitions are intrinsically appealing but they create problems in application because the mode of origin cannot always be determined. Furthermore, evolution is opportunistic and uses whatever genetic information is available, regardless of mode of origin. From the perspective of the evolution of new functions or biological traits in organisms, mode of origin may not be relevant. For these reasons, we deploy a pragmatic definition of novel genes,

meaning genes encoding proteins that are substantially different from, or have no similarity to those in related lineages.

In the present study, our goal was to identify novel genes that originated along the stem lineage of placental mammals. We took advantage of proteome data from twenty vertebrate species and by combining reciprocal BLASTP and MCL clustering were able to identify groups of homologous proteins and determine their relative ages in a phylogenetic context. We generated a total of 20,363 'Homology Groups' (HG), of which 9465 were inferred as present in the common ancestor of placental mammals. The vast majority of these 9465 HG are found more widely that just the placental mammals and therefore date to earlier in metazoan evolution. However, we identified a subset of 357 HG that were present in the most recent common ancestor of the crown Placentalia and are completely absent from all other species (Figure 1, 2). We suggest that these represent genes that arose on the stem lineage of the placental mammals.

Two distinct levels of evolutionary conservation were examined across the 357 HG: (1) Those with moderate to high levels of loss across placental mammals were named Novel Ancestral Placental genes, but each of these was still inferred to have been present in the common ancestor of Placentalia because of retention in representatives of disparate evolutionary lineages; (2) Those HG present in the genomes of all, or all but one, placental mammals in our study (87 HG) were termed Novel Core Placental genes. We suggest that this set of 87 HG represent genes that were central for the emergence of placental mammals, and are involved in biological roles that are highly important for 'being a placental mammal'.

Our analyses suggest that the 357 Novel Ancestral Placental HG are new 'types of genes' that arose on the stem lineage of Placentalia. It is not possible to infer directly the chromosomal location of each new gene at its date of origin, since this would necessitate dating each origin to a time point along a stem lineage that has no living descendants while also knowing the karyotype of each extinct ancestor. As a proxy, we use the human karyotype with the caveat that there have been chromosome fission and fusion events. All but one human chromosome carries Novel Ancestral Placental HG genes, but there is a proportional enrichment on the X and Y chromosomes (Figure 3), known to be homologous across placental mammals with human X chromosome genes also found on the elephant X chromosome [23]. We thus infer that sex chromosomes were a major (but not exclusive) site of origin of the genes on the stem

lineage of placental mammals. Interestingly, the sex chromosomes of placental mammals have a radically different gene composition to those of marsupials (and outgroups) because of a fusion with an autosome bringing new genes to the X chromosome, forming the X Added Region or XAR [24]. We suggest that this event, along with Y chromosome degradation, facilitated the origin of new genes on both sex chromosomes. For both the X and Y, reduced effective population size, lack of recombination, and strong selection in the hemizygous male may have promoted extensive tandem gene duplication and acceleration of DNA sequence evolution.

To gain insight into the contribution that novel genes made to the biology of mammals, we examined gene function and expression using human data. Gene Ontology and KEGG analysis suggested that many Novel Ancestral Placental HG genes have functions in the immune system, in hair and skin development (keratinization), and in the testis. Although these are biological functions known to be complex in mammals as a whole, our analysis focusses specifically on genetic changes on the stem lineage of placental mammals. Hence, if we can safely extrapolate from human data across the placental mammals, we suggest that these functions were subject to extensive evolutionary modification after the divergence of the eutherians from the metatherian and prototherian lineages. This list of functional categories may be incomplete as many human genes within the Novel Ancestral Placental HG have not been assigned a GO term related to a biological process, molecular function or cellular component. This limitation is less extreme for gene expression which we used for an independent insight into gene function, and we were able to examine expression profiles for most genes (Figure 6). As above, this approach highlighted testis as a tissue into which new genes have been recruited and to a lesser extent the immune system. Two additional broad categories of biological function were suggested from human gene expression: functions in the brain and in pre-implantation embryonic development. In each case, many new genes (Novel Ancestral Placental HG) were specifically or predominantly expressed in these RNAseq datasets. Overall, these data suggest there was extensive genetic modification to pathways involved in testis, brain and immune system function and pre-implantation development during eutherian mammal evolution. Almost half of the brain-expressed new genes are on the human X chromosome (19 of 41), consistent with the ‘smart and sexy’ description of the

eutherian X chromosome discussed by Graves [25]. Testes-expressed new genes are found on the human X, Y and autosomes.

An association of new eutherian genes with pre-implantation development has been noted previously, but the current study suggests this is more extensive than formerly recognized and not driven primarily by sex chromosome evolution. For example, several autosomal PRD class homeobox gene families (*ARGFX, DPRX, TPRX, LEUTX, CPHX*) and one autosomal ANTP class homeobox gene (*NANOGNB*) have previously been noted to be specific to placental mammals and expressed in pre-implantation development [17, 26-29]; three of these, *LEUTX*, *CPHX* and *NANOGNB*, were identified in the present study. Additional placental mammal specific genes we identified with enriched expression in preimplantation embryos include: *ZSCAN4*, implicated in pluripotency [30,31] and two members of an extended gene family *KHDC1* and *DPPA5* [32] which have been previously reported as mammal-specific; and a related group of transcriptional repressors, *SSX1-5*, which are frequently over expressed in cancer with reported roles in cell adhesion and migration, cancer stem cell generation and chromatin remodelling [33-36]. These data imply that during the evolution of eutherian mammals there was extensive remodelling of genetic pathways controlling formation of the blastocyst. This conclusion is particularly intriguing in the light of recent embryological work highlighting differences in cell behaviour during the early development of the marsupial Tammar Wallaby compared to placental mammals [4,5]. For example, in human and mouse embryos the early distinction between embryo-fated cells and trophectoderm cells is associated with formation of an inner cell mass within a hollow sphere of cells, while in Tammar wallaby the embryo-fated cells remain as a 'pluriblast' located on the surface of a unilaminar blastocyst layer [4,5]. The functional significance of such differences is not clear, although it is tempting to relate them to the necessity for placental mammals to rapidly establish a distinct and highly active placenta for extended gestation.

Ethics

The authors declare that there are no ethical issues associated with this research.

Data accessibility

Assignment of protein sequences to Homology Groups, size of each Homology Group and processed human gene expression data are uploaded as Electronic Supplementary Material and available at xxxxxxxxxx

The phylogenetic tree data are available under TreeBASE accession http://purl.org/phylo/treebase/phylows/study/TB2:S21443.

Authors' contributions

TLD conceived the study and performed bioinformatic analyses. TLD, PWHH, and JP participated in project design. TLD and PWHH wrote the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013 ERC grant 268513).

Acknowledgements

We thank anonymous reviewers for constructive suggestions.

Footnotes

Electronic Supplementary Material is available online at XXXXXXXXXX

Figures and Tables

Table 1. Genes present in twelve major expression clusters

Figure 1. Taxon sampling and phylogeny. The number of proteins listed for each species is the combined total from NCBI RefSeq and Ensembl protein predictions. Each of the four coloured columns represents a Homology Group. The first two columns are hypothetical examples that would be classified as Novel Ancestral Placental Homology Group, since they contain genes found in one member of the Atlantogenata and one of the Boreoeutheria. The last two columns are hypothetical examples of Novel Core Placental Homology Groups (a subset of Novel Ancestral Placental Homology Groups), being groups found in all, or all but one, placental mammals. 'YES' and 'NO' represent presence or absence of a Homology Group in a species.

Figure 2. BLASTP/MCL pipeline and filtering steps for identifying Novel Ancestral Placental and Novel Core Placental Homology Groups.

Figure 3. Distribution of genes from Novel Ancestral Placental and Novel Core Placental Homology Groups across human chromosomes. The number of proteins in Novel Ancestral Placental and Novel Core Placental Homology Groups are shown per-chromosome as a percentage of the total number of protein coding genes on that chromosome which were present in our dataset. The total number of protein coding genes per-chromosome is plotted on the secondary axis. The significance of the adjusted p-value for the enrichment or depletion of the Novel Ancestral and Novel Core proteins per chromosome are shown in the grid below the histogram (* = p-value < 0.05, ** = p-value < $5e^{-3}$, *** = p-value < $5e^{-29}$).

Figure 4. Phylogenetic tree built using representative proteins from Novel Core Placental Homology Groups. Due to the inherent lack of outgroup the tree was rooted between Atlantogenata and Boreotheria.

Figure 5. GO annotation and pathway enrichment. Genes from Novel Ancestral Placental and Novel Core Placental HG were assessed for enrichment for gene ontology (GO) annotation terms and KEGG pathways. Spot size is proportional to the –log2 of the p-value when a value ≤ 0.05 was found, terms are ordered by significance of enrichment in Novel Ancestral genes. Term and pathways IDs are shown below the term names.

Figure 6. Heatmap of normalised gene expression for 59 human cell types and tissues. Expression data from 59 different human cell types and tissues for 336 different human genes from 249 Novel Ancestral Placental Homology Groups. Clustering is according to expression levels for each gene across all tissues and cell types after normalising each gene's expression to the site of highest expression. Values are shown in a scale between 0 and 1. Individual selected tissue or cell type clusters are labeled on the left edge. The peach colour in the bar running the height of the heatmap identifies those genes which belong to only a Novel Ancestral Placental Homology Groups; a subset are coloured green and identifies those also belonging to a Novel Core Placental Homology Group.

Figure 7. Analysis of clustering and BLASTP results for Novel Core Placental Homology Groups. BLASTP interactions for all proteins within the 87 Novel Core Placental HG were analysed to determine to which, if any, other HG BLASTP hits were detectable. (**A**) BLASTP interactions between the 87 Novel Core Placental HG were assessed to identify which HG had reciprocal BLASTP hits between them. The diagonal line indicates reciprocal hits within an HG to itself. Off-diagonal squares indicate BLASTP interactions between two different Novel Core Placental HG. Black lines illustrate BLASTP interactions between clusters. Numbers 1-5 represent Sets 1-5 in Electronic Supplementary Material Table S6, where more details of the interactions are show. (**B**) BLASTP interactions between the 87 Novel Core Placental HG and all other HG. Black lines between (**A**) and (**B**) are used to illustrate selected examples of where hits were detected. The coloured bars below the plot indicate which species each HG in (B) is present in. A minimum of 25% of the proteins in a Novel Core Placental HG were required to have BLASTP hits against another cluster for a BLASTP interaction to be considered relevant.

Figure 8. Methods of gene evolution. Selected Novel Ancestral Placental Homology Group which contained a single protein were used to examine how selected Homology Groups may have been generated. The syntenic region surrounding the human gene was compared to the equivalent region in opossum. (A) *CCER2* as an example of how a placental mammal protein coding gene has diverged such that it is detected as substantially different to the copy of the gene found in non-placental mammals. (B) Tandem duplication of the *CLPS* loci as an example for how genes can undergo duplication and subsequent divergence, resulting in one or more of the duplicates diverging substantially from the original copy. (C) *IL31* as an example of a gene present in humans but not present in the syntenic location in opossum. (D) Simplified representation of rearrangements surrounding *SPZ1*, as an example of how new genes can be associated with large-scale changes to chromosome structure.

References

1. O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, Giannini NP, Goldberg SL, Kraatz BP, Luo ZX, Meng J, Ni X, Novacek MJ, Perini FA, Randall ZS, Rougier GW, Sargis EJ, Silcox MT, Simmons NB, Spaulding M, Velazco PM, Weksler M, Wible JR, Cirranello AL. The placental mammal ancestor and the post-K-Pg radiation of placentals. Science 2013; 339: 662-7.
2. Dos Reis M, Donoghue PCJ, Yang Z. Neither phylogenomic nor palaeontological data support a Palaeogene origin of placental mammals. Biol Lett 2014; 10:20131003.
3. Moffett A, Loke C. Immunology of placentation in eutherian mammals. Nat Rev Immunol. 2006; 6: 584-94.
4. Rossant J, Tam PP. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development 2009; 136: 701-13.
5. Frankenberg S, Shaw G, Freyer C, Pask AJ, Renfree MB. Early cell lineage specification in a marsupial: a case for diverse mechanisms among mammals. Development 2013; 140: 965-75.
6. Lavialle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, Heidmann T. Paleovirology of '*syncytins*', retroviral *env* genes exapted for a role in placentation. Philos Trans R Soc Lond B. 2013; 368: 20120507.
7. Holland PWH, Marlétaz F, Maeso I, Dunwell TL, Paps J. New genes from old: asymmetric divergence of gene duplicates and the evolution of development. Philos Trans R Soc Lond B. 2017; 372: 20150480.
8. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421.
9. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002; 30: 1575-84.
10. Dunwell TL. GitHub, 2017. doi: 10.6084/m9.figshare.5340778
11. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30: 772-80.
12. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009; 25: 1972-3.

13. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010; 5: e9490.
14. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics. 2009; 25: 2286-8.
15. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4: 44-57.
16. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protocols 2012;* 7: 562-578.
17. Dunwell TL, Holland PWH. Diversity of human and mouse homeobox gene expression in development and adult tissues. BMC Dev Biol. 2016; 16: 40.
18. Paps J, Holland PWH. What makes an animal? Reconstruction of the ancestral metazoan genome reveals an explosion of novelty. In review.
19. Morgan CC, Foster PG, Webb AE, Pisani D, McInerney JO, O'Connell MJ. Heterogeneous models place the root of the placental mammal phylogeny. Mol Biol Evol. 2013; 30: 2145-56.
20. Kordis D, Turk V. Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes. BMC Evol Biol. 2009; 9: 266.
21. Papenfuss AT, Feng ZP, Krasnec K, Deakin JE, Baker ML, Miller RD. Marsupials and monotremes possess a novel family of MHC class I genes that is lost from the eutherian lineage. BMC Genomics. 2015; 16: 535.
22. Neme R, Tautz D. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. BMC Genomics. 2013; 14:117.
23. Delgado CL, Waters PD, Gilbert C, Robinson TJ, Graves JA. Physical mapping of the elephant X chromosome: conservation of gene order over 105 million years. Chromosome Res. 2009; 17: 917-26.
24. Graves JA. Did sex chromosome turnover promote divergence of the major mammal groups? De novo sex chromosomes and drastic rearrangements may have posed reproductive barriers between monotremes, marsupials and placental mammals. BioEssays. 2016; 38: 734-43.

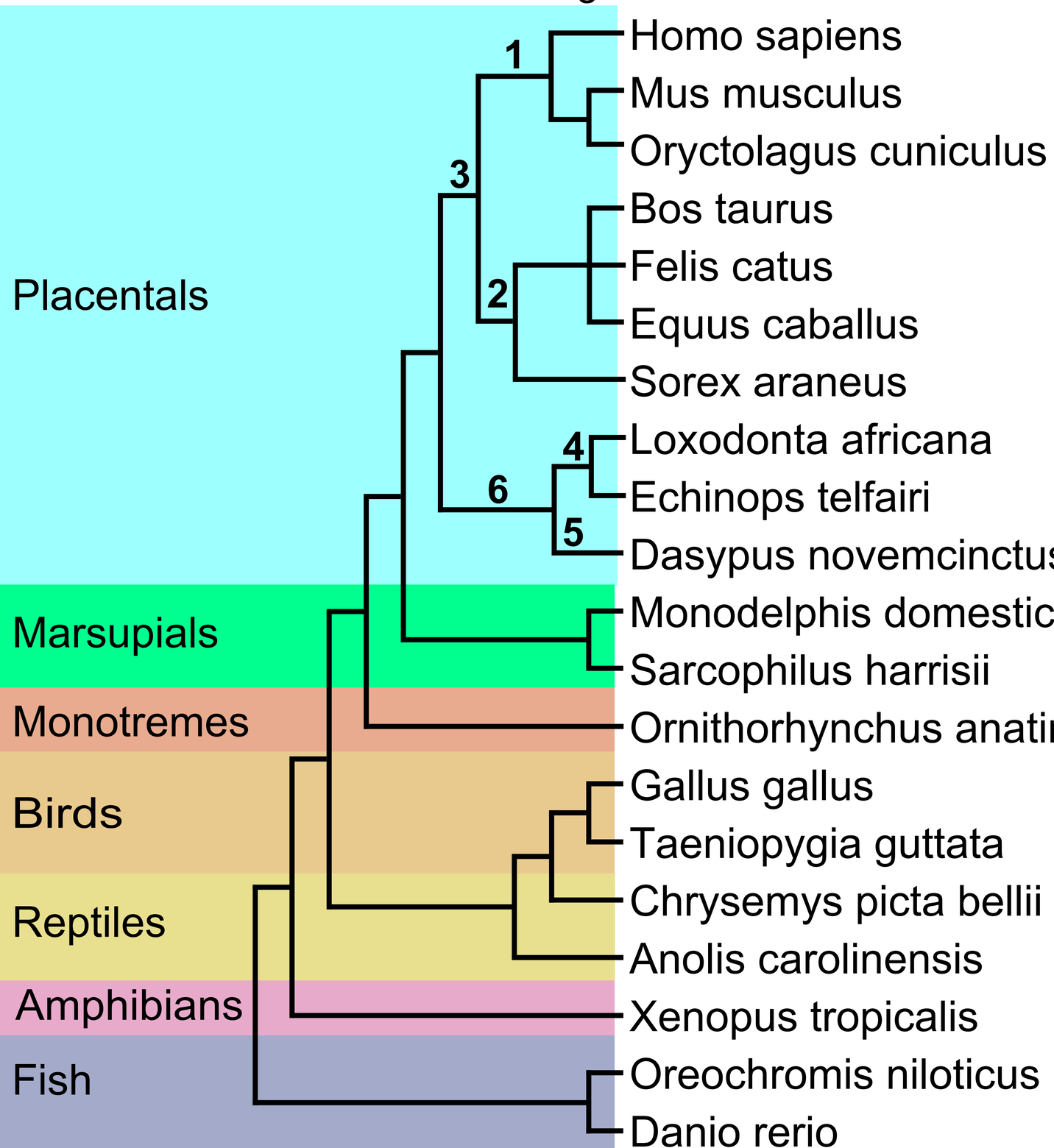
25. Graves JAM, Gecz J, Hameister H. Evolution of the human X – a smart and sexy chromosome that controls speciation and development. 2002; Cytogenet Genome Rees 99: 141-145.
26. Töhönen V, Katayama S, Vesterlund L, Jouhilahti EM, Sheikhi M, Madissoon E, Filippini-Cattaneo G, Jaconi M, Johnsson A, Bürglin TR, Linnarsson S, ovatta O, Kere J. Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. Nat Commun. 2015; 6: 8207.
27. Maeso I, Dunwell TL, Wyatt CD, Marlétaz F, Vető B, Bernal JA, Quah S, Irimia M, Holland PW. Evolutionary origin and functional divergence of totipotent cell homeobox genes in eutherian mammals. BMC Biol. 2016; 14: 45.
28. Madissoon E, Jouhilahti EM, Vesterlund L, Töhönen V, Krjutškov K, Petropoulous S, Einarsdottir E, Linnarsson S, Lanner F, Månsson R, Hovatta O, Bürglin TR, Katayama S, Kere J. Characterization and target genes of nine human PRD-like homeobox domain genes expressed exclusively in early embryos. Sci Rep. 2016; 6: 28995.
29. Dunwell TL, Holland PWH. A sister of NANOG regulates genes expressed in pre-implantation human development. Open Biology. 2017; 7: 170027.
30. Hirata T, Amano T, Nakatake Y, Amano M, Piao Y, Hoang HG, Ko MS. Zscan4 transiently reactivates early embryonic genes during the generation of induced pluripotent stem cells. Sci Rep. 2012; 2: 208.
31. Jiang J, Lv W, Ye X, Wang L, Zhang M, Yang H, Okuka M, Zhou C, Zhang X, Liu L, et al. Zscan4 promotes genomic stability during reprogramming and dramatically improves the quality of iPS cells as demonstrated by tetraploid complementation. Cell Res. 2013; 23: 92–106.
32. Pierre A, Gautier M, Callebaut I, Bontoux M, Jeanpierre E, Pontarotti P, Monget P. Atypical structure and phylogenomic evolution of the new eutherian oocyte- and embryo-expressed KHDC1/DPPA5/ECAT1/OOEP gene family. Genomics 2007; 90: 583-94.
33. Cronwright G, Le Blanc K, Götherström C, Darcy P, Ehnman M, Brodin B. Cancer/testis antigen expression in human mesenchymal stem cells: down-regulation of SSX impairs cell migration and matrix metalloproteinase 2 expression. Cancer Res. 2005; 65: 2207-15.

559 34. Yang P, Huo Z, Liao H, Zhou Q. Cancer/testis antigens trigger epithelial-mesenchymal
560 transition and genesis of cancer stem-like cells. Curr Pharm Des. 2015; 21: 1292-300
561 35. Zöllner SK, Rössig C, Toretsky JA. Synovial sarcoma is a gateway to the role of
562 chromatin remodeling in cancer. Cancer Metastasis Rev. 2015; 34: 417-28.
563 36. Bloom JE, McNeel DG. SSX2 regulates focal adhesion but does not drive the epithelial
564 to mesenchymal transition in prostate cancer. Oncotarget. 2016; 7: 50997-1011.
565

Tissue(s)	Gene(s)
All or many	ARMCX5, C22orf29, DCAF16, EID1, EID2, FAM127C, FAM156A, FAM156B, HMGN1, MRFAP1L1, NBPF1, NBPF11, NBPF12, NBPF14, NBPF15, NBPF19, NBPF26, NBPF8, NBPF9, RBM3, RPL41, TCEAL1, TCEAL4
Brain	ARMCX4, ARMCX5-GPRASP, BEX1, BEX2, BEX4, BEX5, BHLHB9, C12orf76, C1orf122, C22orf24, C6orf1, CASC10, DEXI, EID2B, ENHO, FAM127A, FAM127B, GPRASP1, GPRASP2, HEPN1, IGIP, IRGQ, LDOC1, LDOC1L, LOC728392, MRFAP1, PCSK1N, PNMAL2, RGAG4, RNF187, RPT5, RTP5, SMIM17, SNURF, TCEAL2, TCEAL3, TCEAL5, TCEAL6, TCEAL7, TMEM155, TMEM88B
Breast	CLPSL2, CSN1S1, CSN2, CSN3, SCGB2A2
ESC	ADM5, C10orf111, DEFB107B, NGFRAP1, LOC105377021, TCEAL8
Fallopian Tube	CCDC114, CCHC12, SCGB2A1, SMIM6
Immune cells	ANXA2R, ARIH2OS, LOC100128108, LOC101929599, LOC105371437, LOC105372412, NCR3, PVRIG, SECTM1
Late Blastocyst	GNRH2, LOC101928585, LUZP6, RUSC1-AS1
Oocyte, Zygote, 2-Cell and 4-Cell embryo	C10orf95, C3orf56, CXorf67, DPPA5, ERICH5, FAM24B, GML, HJURP, LOC105371430, PRR32, SCGB2B2, WFDC10A
Salivary Gland	MUC7, PROL1, PRR4, PRR27, SMR3A, SMR3B
Testes	BPIFA3, C10orf55, C11orf71, C12orf42, C16orf82, C17orf112, C1orf105, C20orf141, C20orf173, C6orf10, C7orf61, C9orf50, CABS1, CCDC179, CPXCR1, CSTL9, CT62, CXorf66, CYLC1, DEFB119, DEFB121, DEFB123, FAM24A, HMGN5, HSPB9, INSL6, KIAA1210, LOC100505478, LOC100506217, LOC730183, LYPD4, NBPF3, NBPF6, NPAP1, PAGE3, PRM2, PROCA1, RNASE11, SBPF4, SIGLECL1, SMCP, SMIM2, SPATA12, SPATA3, SPATA32, SPZ1, TEX22, TMEM191B, TMEM191C, TMEM31, TNP2, TRPC5OS, TSPY1, TSPY10, TSPY2, TSPY3, TSPY4, TSPY8, UBE2Q2L, ULBP1, ULBP3
8-Cell	CT47A1, CT47A10, CT47A11, CT47A12, CT47A2, CT47A3, CT47A4, CT47A5, CT47A6, CT47A7, CT47A8, CT47A9, LEUTX, LOC105373368
8-Cell, Morula	BAGE2, BIK, CSAG1, CSAG2, CSAG3, CT47B1, CXorf49B, CXorg49, DEFB124, KHCD1, KHDC1L, LOC101059915, LOC102724657, LOC105371346, LUZP4, NANOGNB, PRR23A, PRR23B, PRR23C, SSX1, SSX2, SSX2B, SSX3, SSX4, SSX5, SXX4B, TEX19, WBP5, XAGE5, ZNF576, ZSCAN4

Clades within placental mammals

1 - Euarchontoglires
2 - Laurasiatheria
3 - Boreotheria
4 - Afrotheria
5 - Xenarthra
6 - Atlantogenata

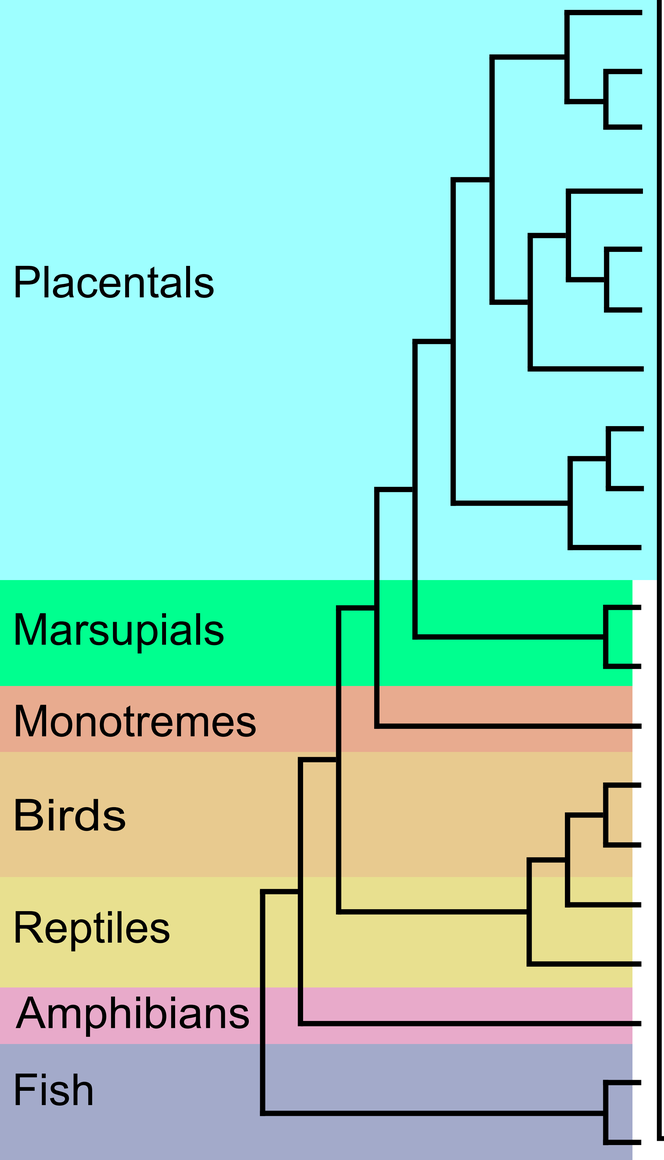


	Proteins in data set	Novel Ancestral		Novel Core	
		OR		OR	
Homo sapiens	21178	NO	NO	YES	YES
Mus musculus	23983	NO	NO	YES	YES
Oryctolagus cuniculus	23113	YES	NO	YES	YES
Bos taurus	24534	NO	NO	YES	NO
Felis catus	21481	NO	NO	YES	YES
Equus caballus	22458	NO	YES	YES	YES
Sorex araneus	23896	NO	NO	YES	YES
Loxodonta africana	22445	NO	NO	YES	YES
Echinops telfairi	25021	YES	NO	YES	YES
Dasypus novemcinctus	26265	NO	YES	YES	YES
Monodelphis domestica	24090	NO	NO	NO	NO
Sarcophilus harrisii	22059	NO	NO	NO	NO
Ornithorhynchus anatinus	26365	NO	NO	NO	NO
Gallus gallus	18537	NO	NO	NO	NO
Taeniopygia guttata	19556	NO	NO	NO	NO
Chrysemys picta bellii	21194	NO	NO	NO	NO
Anolis carolinensis	21641	NO	NO	NO	NO
Xenopus tropicalis	24262	NO	NO	NO	NO
Oreochromis niloticus	27256	NO	NO	NO	NO
Danio rerio	28964	NO	NO	NO	NO

468,298 individual peptides obtained from NCBI and Ensembl.

219,303,016,804 reciprocal blast comparisons.

182,343,954 hits (0.08%) equal to or below evalue 5e-5



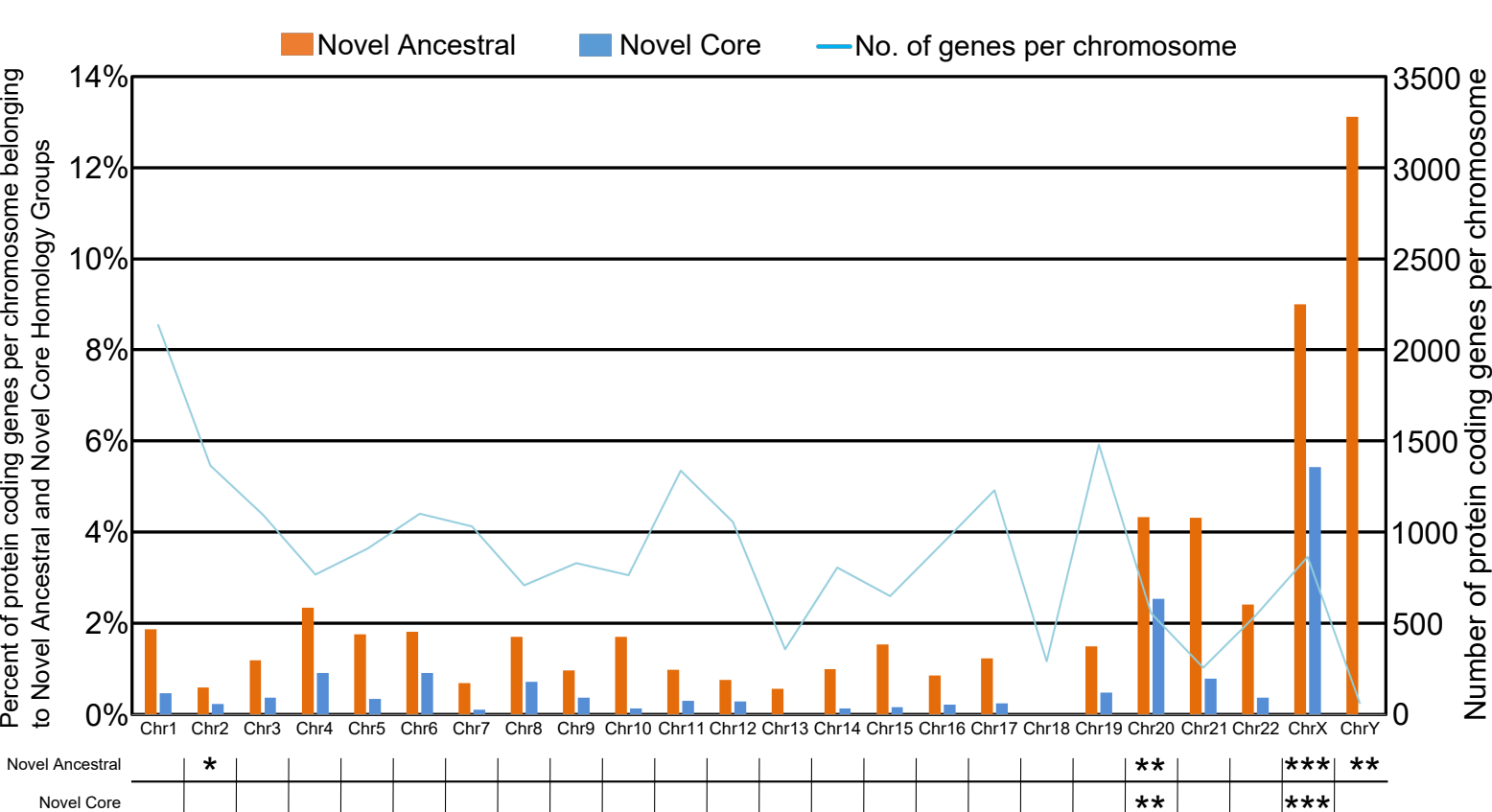
20,363 Homology Groups generated by MCL.

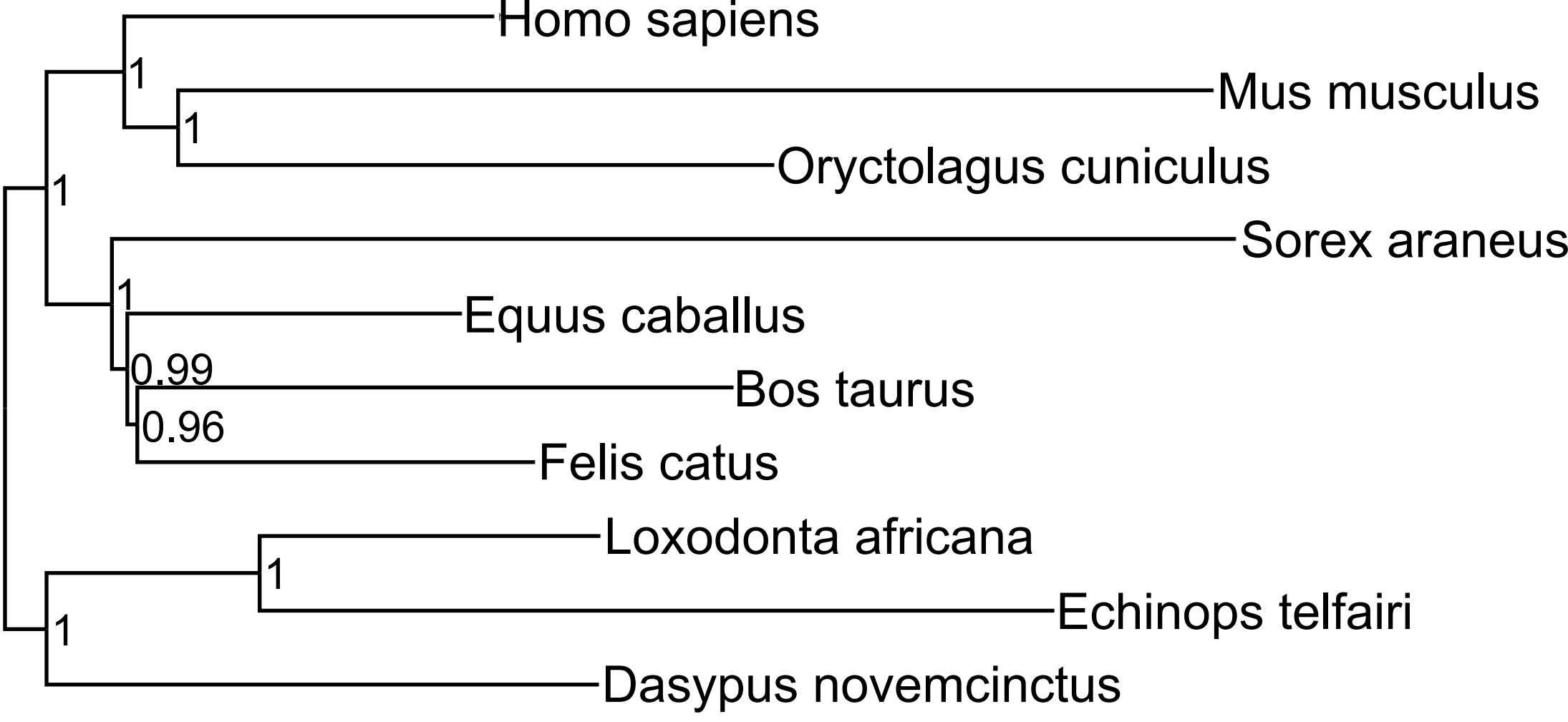
9465 Homology Groups present in the common ancestor of placental mammals.

357 'Novel Ancestral' Homology Groups present in the common ancestor of placental mammals. Of these, 249 clusters containing 376 proteins are present in humans.

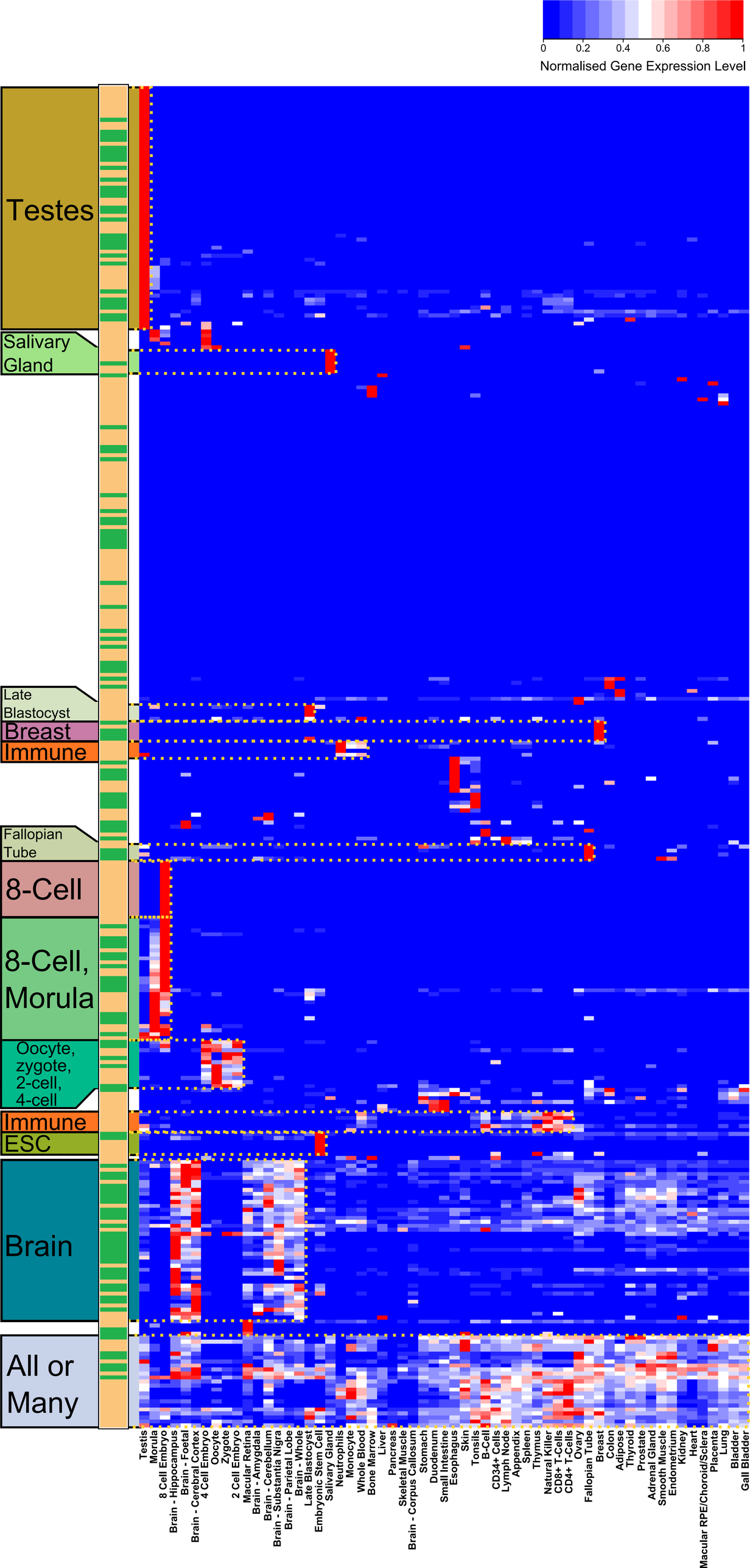
87 'Novel Core' Homology Groups present in all (or all but one) placental mammals. Of these, 86 are present in humans containing 133 proteins.

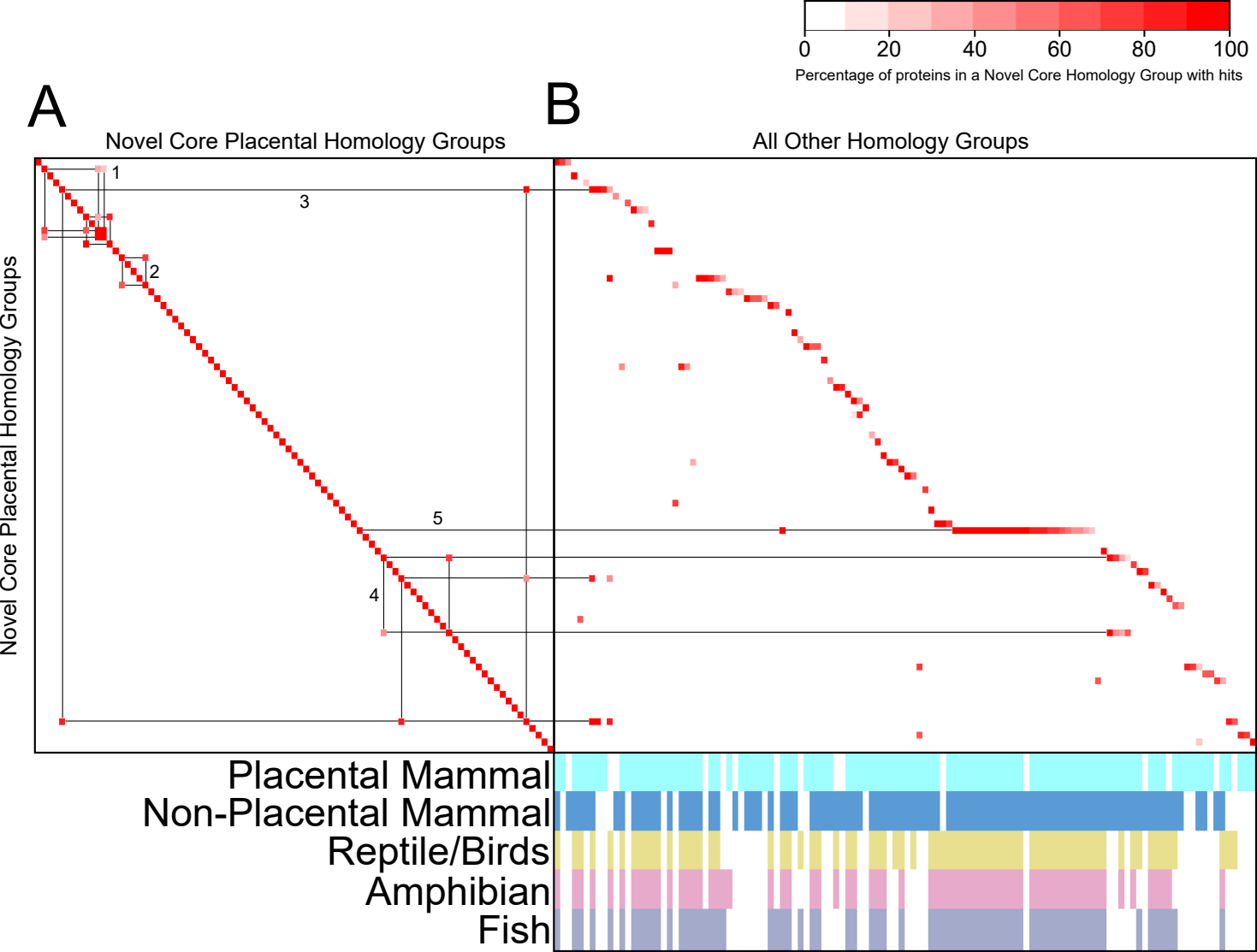
Proteins in these clusters are central for making a placental mammal.


Novel Ancestral
Novel Core
No. of genes per chromosome
Percent of protein coding genes per chromosome belonging to Novel Ancestral and Novel Core Homology Groups
Number of protein coding genes per chromosome
14%
12%
10%
8%
6%
4%
2%
0%
3500
3000
2500
2000
1500
1000
500
0
Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12 Chr13 Chr14 Chr15 Chr16 Chr17 Chr18 Chr19 Chr20 Chr21 Chr22 ChrX ChrY
Novel Ancestral
Novel Core
*
**
***
**
**
***

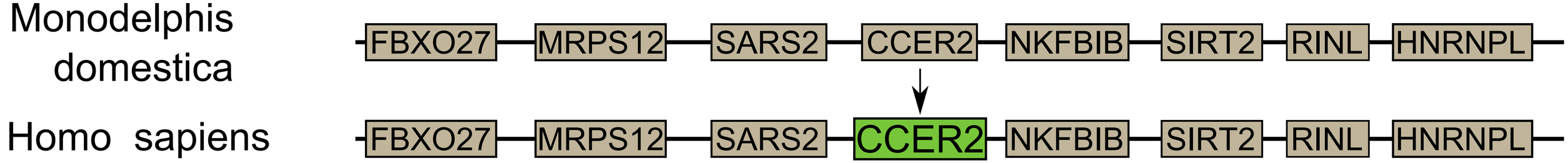

Homo sapiens
Mus musculus
Oryctolagus cuniculus
Sorex araneus
Equus caballus
Bos taurus
Felis catus
Loxodonta africana
Echinops telfairi
Dasypus novemcinctus
1
1
1
1
1
0.99
0.96
1

		Novel Ancestral	Novel Core
Biological Process	Defense response to bacterium GO:0042742		
	Gonadal mesoderm development GO:0007506		
	Keratinization GO:0031424		
	Peptide cross-linking GO:0018149		
	Negative regulation of nucleic acid-templated transcription GO:1903507		
	Innate immune response GO:0045087		
	Keratinocyte differentiation GO:0030216		
	Natural killer cell mediated cytotoxicity GO:0042267		
	Spermatogenesis GO:0007283		
	Antigen processing and presentation GO:0019882		
	Cell differentiation GO:0030154		
	Regulation of transcription from RNA polymerase II promoter GO:0006357		
	Transcription, DNA-templated GO:0006351		
	Natural killer cell activation GO:0030101		
Cellular Component	Extracellular region GO:0005576		
	Cornified envelope GO:0001533		
	Cell surface GO:0009986		
Molecular Function	WW domain binding GO:0050699		
	Natural killer cell lectin-like receptor binding GO:0046703		
	Structural molecule activity GO:0005198		
	Transcription corepressor activity GO:0003714		
	Antigen binding GO:0003823		
KEGG Pathway	Natural killer cell mediated cytotoxicity hsa04650		
	Transcriptional misregulation in cancer hsa05202		

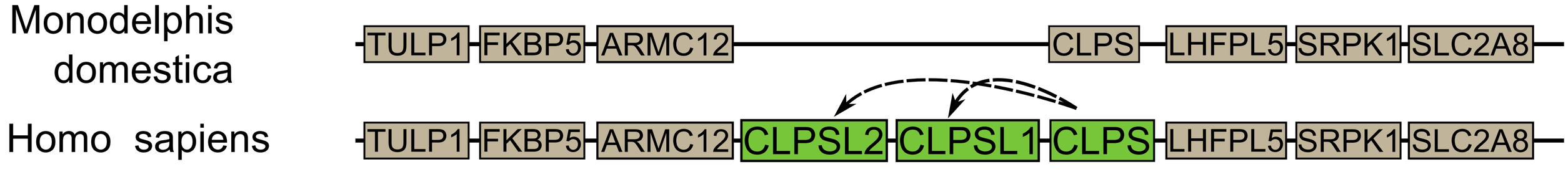

0
0.2
0.4
0.6
0.8
1
Normalised Gene Expression Level
Testes
Salivary Gland
Late Blastocyst
Breast
Immune
Fallopian Tube
8-Cell
8-Cell, Morula
Oocyte, zygote, 2-cell, 4-cell
Immune
ESC
Brain
All or Many
Testis
Morula
8 Cell Embryo
Brain - Hippocampus
Brain - Foetal
Brain - Cerebral Cortex
4 Cell Embryo
Oocyte
Zygote
2 Cell Embryo
Macular Retina
Brain - Amygdala
Brain - Cerebellum
Brain - Substantia Nigra
Brain - Parietal Lobe
Brain - Whole
Late Blastocyst
Embryonic Stem Cell
Salivary Gland
Neutrophils
Monocyte
Whole Blood
Bone Marrow
Liver
Pancreas
Skeletal Muscle
Brain - Corpus Callosum
Stomach
Duodenum
Small Intestine
Esophagus
Skin
Tonsils
B-Cell
CD34+ Cells
Lymph Node
Appendix
Spleen
Thymus
Natural Killer
CD8+ T-Cells
CD4+ T-Cells
Ovary
Fallopian Tube
Breast
Colon
Adipose
Thyroid
Prostate
Adrenal Gland
Smooth Muscle
Endometrium
Kidney
Heart
Macular RPE/Choroid/Sclera
Placenta
Lung
Bladder
Gall Bladder


0
20
40
60
80
100
Percentage of proteins in a Novel Core Homology Group with hits
A
Novel Core Placental Homology Groups
B
All Other Homology Groups
Novel Core Placental Homology Groups
1
3
2
5
4
Placental Mammal
Non-Placental Mammal
Reptile/Birds
Amphibian
Fish

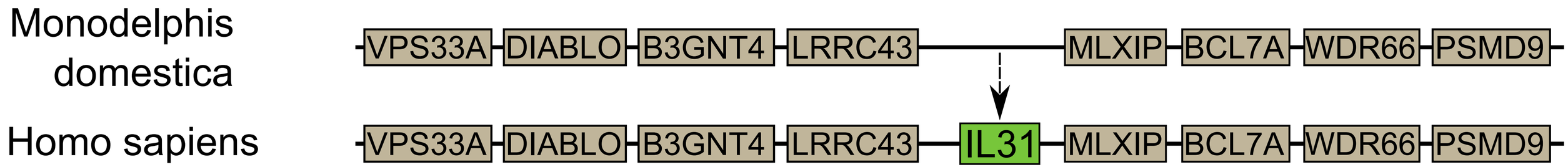
A) Divergence of an established gene



B) Tandem duplication and divergence of a gene



C) Appearance of 'de novo' coding sequence



D) Association with chromosomal break points and/or rearrangements

Novel and divergent genes in the evolution of placental mammals

Dunwell TL, Paps J, Holland PWH

Legends for Electronic Supplemental Material

Figure S1. Heatmap of normalised human gene expression showing gene names

Same data and analysis as Figure 6 but showing gene names.

Table S1. **Protein sequence accession numbers**

List of NCBI and Ensembl protein IDs used to generate the combined data set, numerical identifiers for the Homology Group each protein was placed into, and indication of whether genes/HG were assigned to Novel Ancestral Placental and Novel Core Placental HG. Excel file.

Table S2. **Numbers of proteins analysed per species**

The number of protein IDs in the original NCBI and Ensembl protein data used. Excel file.

Table S3. **Assignment of proteins to Homology Groups**

List of all 20363 Homology Groups giving the number of proteins in each Homology Group in each species, and which HG belong to the Novel Ancestral Placental and Novel Core Placental categories. Excel file.

Table S4. Proteins used for phylogenetic analysis

IDs of the selected proteins from each Novel Core Placental Homology Group used for phylogenetic analysis, including amino acid sequences after alignment and trimming. Excel file.

Table S5. **Expression data for human genes**

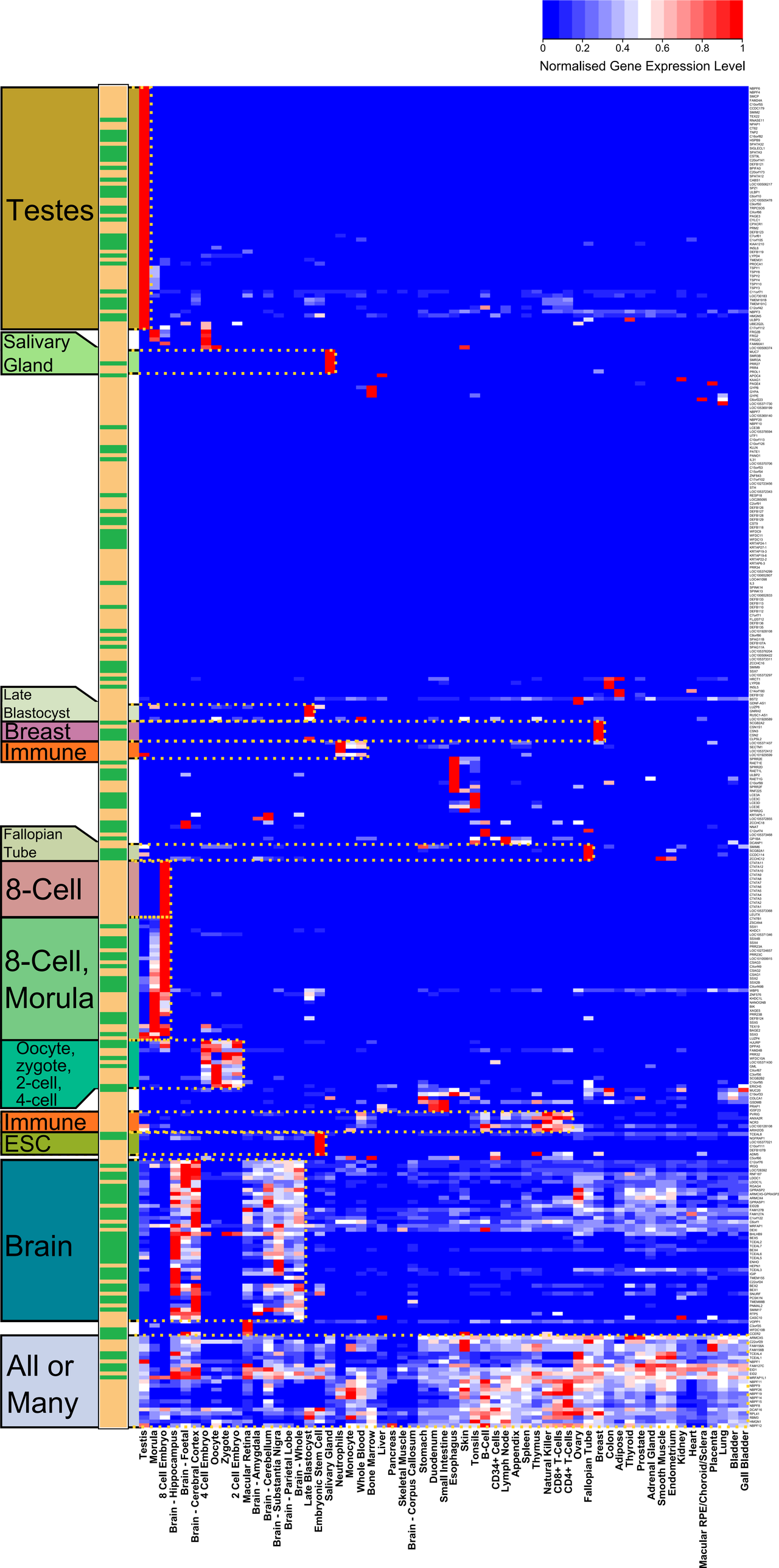
Raw and normalised FPKM gene expression values for all human genes in Novel Ancestral Placental and Novel Core Placental Homology Groups. Excel file.

Table S6. **Examples of sequence similarity searches using Novel Core Placental Homology Groups**

Details of BLASTP cluster interactions (1-5) highlighted in Figure 7A.

Table S7. **Sequence similarity searches for all Novel Core Placental Homology Groups**

Details of BLASTP cluster interactions between Novel Core and all other homology groups, as shown in Figure 7.


0
0.2
0.4
0.6
0.8
1
Normalised Gene Expression Level
Testes
Salivary Gland
Late Blastocyst
Breast
Immune
Fallopian Tube
8-Cell
8-Cell, Morula
Oocyte, zygote, 2-cell, 4-cell
Immune
ESC
Brain
All or Many
Testis
Morula
8 Cell Embryo
Brain - Hippocampus
Brain - Foetal
Brain - Cerebral Cortex
4 Cell Embryo
Oocyte
Zygote
2 Cell Embryo
Macular Retina
Brain - Amygdala
Brain - Cerebellum
Brain - Substantia Nigra
Brain - Parietal Lobe
Brain - Whole
Late Blastocyst
Embryonic Stem Cell
Salivary Gland
Neutrophils
Monocyte
Whole Blood
Bone Marrow
Liver
Pancreas
Skeletal Muscle
Brain - Corpus Callosum
Stomach
Duodenum
Small Intestine
Esophagus
Skin
Tonsils
B-Cell
CD34+ Cells
Lymph Node
Appendix
Spleen
Thymus
Natural Killer
CD8+ T-Cells
CD4+ T-Cells
Ovary
Fallopian Tube
Breast
Colon
Adipose
Thyroid
Prostate
Adrenal Gland
Smooth Muscle
Endometrium
Kidney
Heart
Macular RPE/Choroid/Sclera
Placenta
Lung
Bladder
Gall Bladder