

# Baby Genomics: Tracing the Evolutionary Changes That Gave Rise to Placentation

Yue Hao <sup>1,\*</sup>, Hyuk Jin Lee<sup>2</sup>, Michael Baraboo<sup>3</sup>, Katherine Burch<sup>4</sup>, Taylor Maurer<sup>5</sup>, Jason A. Somarelli<sup>6,7</sup>, and Gavin C. Conant<sup>1,8,9,10</sup>

<sup>1</sup>Bioinformatics Research Center, North Carolina State University

<sup>2</sup>Division of Biological Sciences, University of Missouri-Columbia

<sup>3</sup>Department of Computer Science, Truman State University

<sup>4</sup>Department of Psychology, Truman State University

<sup>5</sup>Department of Biology, Kenyon College

<sup>6</sup>Duke Cancer Institute, Duke University Medical Center

<sup>7</sup>Department of Medicine, Duke University School of Medicine

<sup>8</sup>Division of Animal Sciences, University of Missouri-Columbia

<sup>9</sup>Program in Genetics, North Carolina State University

<sup>10</sup>Department of Biological Sciences, North Carolina State University

\*Corresponding author: E-mail: yhao7@ncsu.edu.

Accepted: February 2, 2020

## Abstract

It has long been challenging to uncover the molecular mechanisms behind striking morphological innovations such as mammalian pregnancy. We studied the power of a robust comparative orthology pipeline based on gene synteny to address such problems. We inferred orthology relations between human genes and genes from each of 43 other vertebrate genomes, resulting in ~18,000 orthologous pairs for each genome comparison. By identifying genes that first appear coincident with origin of the placental mammals, we hypothesized that we would define a subset of the genome enriched for genes that played a role in placental evolution. We thus pinpointed orthologs that appeared before and after the divergence of eutherian mammals from marsupials. Reinforcing previous work, we found instead that much of the genetic toolkit of mammalian pregnancy evolved through the repurposing of preexisting genes to new roles. These genes acquired regulatory controls for their novel roles from a group of regulatory genes, many of which did in fact originate at the appearance of the eutherians. Thus, orthologs appearing at the origin of the eutherians are enriched in functions such as transcriptional regulation by Krüppel-associated box-zinc-finger proteins, innate immune responses, keratinization, and the melanoma-associated antigen protein class. Because the cellular mechanisms of invasive placentae are similar to those of metastatic cancers, we then used our orthology inferences to explore the association between placenta invasion and cancer metastasis. Again echoing previous work, we find that genes that are phylogenetically older are more likely to be implicated in cancer development.

**Key words:** comparative genomics, orthology inference, mammalian pregnancy, placental mammals.

## Introduction

Comparative genomics is a powerful tool that allows researchers to infer the genetic changes underlying evolutionary transitions: Correctly inferring the phylogenetic point of origin of the genes involved in those transitions is crucial for that purpose. We illustrate this idea here by considering two important and related biological processes in mammals:

placental formation during pregnancy and the exploitation of host genes in cancer development (Kshitiz et al. 2019), showing how a consideration of the phylogenetic history of the genes involved in them is enlightening. In the first example, the evolutionary origin of mammalian pregnancy is an intriguing process involving reprogramming of ancient gene regulatory networks (Lynch et al. 2015), as well as innovations

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

such as the emergence of new cell types (Wagner et al. 2014; Chavan et al. 2016) and the modulation of the immune system to allow for the internal maintenance of an organism genetically distinct from the mother (Mincheva-Nilsson and Baranov 2014). Identifying the group of novel genes that originated at the radiation of the placental mammals would shed light on how reproductive strategies were changed in those mammals by genetic pathway remodeling (Dunwell et al. 2017).

Viviparity (internal embryo development and live birth) has evolved multiple times in vertebrates, including in bony and cartilaginous fish, squamates, and mammals (Van Dyke et al. 2014). However, the degree of maternal–fetal tissue integration varies significantly across these groups (Roberts et al. 2016). The extant mammals contain two subclasses: prototheria, which contains the order monotremata (e.g., platypus), and theria, which contains the metatheria (marsupial) and eutheria (placental) clades (Smith 2015). Eutherian mammals diverged from marsupials about 160 Ma (Meredith et al. 2011; dos Reis et al. 2012; O’Leary et al. 2013; Tarver et al. 2016; Wu et al. 2017). Among therian mammals, marsupials and eutherians are both viviparous, and the extraembryonic membrane in marsupials is homologous to the placenta in eutherian mammals (Wildman 2016). However, the two groups have evolved differing reproductive strategies: marsupials change their milk composition dramatically during the extended lactation window in compensation for their more primitive, underdeveloped, placentae (and hence offspring at birth), whereas eutherians acquired more complicated placentae, allowing the offspring to be more developed at birth (Lefèvre et al. 2010; Guernsey et al. 2017).

Mammalian placentae are formed by close apposition of maternal and fetal tissues (Mossman 1937) and have the vital functions of exchanging nutrients, dissolved gases, and waste during fetal development. Placental morphology and the invasiveness of fetal tissue (how much it erodes the maternal uterine lining) varies greatly across eutherian mammals (Smith 2015). The least invasive are the epitheliochorial placentae (e.g., in pigs, cows, and horses), whereas the endotheliochorial placentae (e.g., in cats, bats, and elephants) invade by erosion of the trophoblast in uterine lining. The most invasive form are the hemochorial placentae (e.g., in mice and humans), with deeper fusion of the fetal tissue into the maternal tissue (Mossman 1987; Wildman et al. 2006). Notably, the invasive hemochorial placenta has been reconstructed to be the ancestral state of the eutherian placenta, based on its phylogenetic distribution (Vogel 2005; Wildman et al. 2006).

The evolution of mammalian pregnancy at the genomic level was complex. Rather than a burst of gene evolution at the origin of the placental mammals, a large set of genes contributing to placentation already existed, but their functions shifted or expanded during the origin of eutherian mammals (Guernsey et al. 2017), probably through a large-scale rewiring of gene regulatory networks that was at least partly

mediated by the insertion and repurposing of transposable elements (TEs) and retroviruses (Kriegs et al. 2006; Churakov et al. 2009; Lynch et al. 2011). This hypothesis is supported by the observation that many of the genes expressed early in mouse placental development are evolutionarily much older than the mammalian radiation, whereas genes expressed later showed a higher propensity to be rodent-specific (Knox and Baker 2008). The ancient retroviruses, relics of whose genomes were exploited during the evolutionary origin of placental mammals, presumably also predate the eutherian divergence (Lee et al. 2013). As a result, many genes and sequences that were ancestrally expressed in other organs were recruited to be expressed in the endometrium and gained new functions related to maternal–fetal interactions (Lynch et al. 2015). A novel type of cell, decidual stromal cells, also appeared along the stem branch of eutherian mammals (Wagner et al. 2014). Among the functions of these new cells is thought to be a role in preventing immune rejection and the inflammation caused by maternal–fetal conflict (Chavan et al. 2016). It even appears that part of this immune response has been repurposed to facilitate placenta formation: during trophoblast invasion, an innate immune response by natural killer cells is needed for proper placentation (Faas and de Vos 2017, 2018).

The evolutionary origins of placentation are of interest for many reasons, including that of what it can tell us about the functions of the immune system and the nature of certain diseases. For instance, the invasion of trophoblast tissue into the maternal tissue incurs an innate immune response, whereas eutherian mammals have evolved mechanisms to repress maternal–fetal immune conflict to enable viviparity (Faas and de Vos 2017, 2018). Similarly, our other example of using the evolutionary age of genes for biological exploration comes from the analogy between placentation and metastasis (Kshitiz et al. 2019). Because these two processes have similarities at the cellular level, placentation genes may also be overrepresented in genes that are apt to change expression due to cancer. Perhaps strikingly, malignant tumors are suppressed in mammalian lineages that have evolved non-invasive placentae (D’Souza and Wagner 2014; Stearns and Medzhitov 2016). It is also known that phylogenetically older genes tend to be associated with human diseases more generally (Domazet-Lošo and Tautz 2008) and that one pathway for cancer development can be mutations that disrupt the communication between new and old genes (Trigos et al. 2019). We have therefore identified a list of genes that are involved in both placenta invasiveness and in cancer metastasis and inferred their evolutionary ages on a rather finer time-scale than the previous analyses (Ferretti et al. 2007; Holtan et al. 2009; D’Souza and Wagner 2014).

A reliable method for estimating gene age is to infer orthologous genes across species on a known phylogeny, that is, to find homologous genes that last shared a common ancestor at a speciation event (Koonin 2005). Orthology relations are

often used to infer evolutionary histories (Kristensen et al. 2011) and to provide functional annotation for uncharacterized homologs (Kriventseva et al. 2015), as such genes tend to conserve their molecular and biological functions (Koonin and Galperin 2003). Many common orthology detection methods are based on sequence similarity. For example, OrthoDB (Kriventseva et al. 2015) first identifies a set of putative orthologs on the basis of reciprocal best hits (Ward and Moreno-Hagelsieb 2014) which can rapidly identify pairs of orthologs between two species. It then uses hierarchical clustering of orthologs to refine the search (Kriventseva et al. 2008). We have previously developed a pipeline that adds the use of gene order information (e.g., synteny) to data on sequence similarity in order to infer orthology relationships between pairs of genomes (Conant 2009; Bekaert and Conant 2011, 2014). When these pairwise estimates are placed on a phylogeny, they allow us to estimate relative gene ages. Because this use of synteny distinguishes our pipeline from most existing orthology inference tools (Zerbino et al. 2018) and pregnancy-related gene prediction tools (Kim et al. 2016) and because it resolves more multigene families into one-to-one orthologous relations than these methods, we used a rewritten version of it to explore two questions: 1) Understanding the genomic changes underlying the genetic origins of mammalian pregnancy and tracing how the gene “toolkit” for placentation evolved and 2) understanding how placental-related genes and mammal orthologs more generally can be used to infer whether a gene has a propensity to be modified or exploited in cancer development.

Our results show that most genes expressed in placenta have evolutionary origins prior to the eutherian radiation, suggesting that they were repurposed or gained new functions in this evolutionary transition. We found that orthologs shared by eutherian ancestor but not any earlier ancestor are enriched in functions and protein classes such as transcription regulation by Krüppel-associated box (KRAB)-zinc-finger proteins (ZNFs), natural killer cell activity, keratinization, and the Melanoma Antigen Gene (MAGE) protein family, indicating that there were large-scale changes in transcriptional regulation and immune response at the origin of mammalian pregnancy. As expected from studies with coarser timescales, genes implicated in tumor development tend to be evolutionarily older, even within the relatively restricted timeframe of the mammalian radiation.

## Materials and Methods

### Genomic Data

The complete set of coding sequences from the genomes of 44 amniotes (3 birds, 1 monotreme, 3 marsupials, and 37 eutherian mammals) was downloaded from the Ensembl database (release 84, Zerbino et al. 2018). From these

genomes, we extracted the longest transcript of each protein-coding gene.

### Orthology Assignment Pipeline

Using these data, we sequentially compared each other genome with the human genome using our pipeline for orthology inference: ORthology Inference using Synteny (ORIS; Conant 2009; Bekaert and Conant 2011, 2014). Since our previous analyses, we have improved ORIS both by dramatically expanding the set of genomes searched (44 versus 18) and by improving the sensitivity of the homology search with a multipass approach using GenomeHistory (Conant and Wagner 2002; [supplementary table S1, Supplementary Material](#) online). We also included three bird genomes in our analysis to allow better basal representation at the root branch of the phylogenetic tree that we used for tracing the origins of each orthologous group. By varying the sequence similarity required for a match, as well as the seed parameters used when deciding to spawn alignments, we have improved the pipeline’s performance when used with genomes more distantly related to humans. The parameter values used for the searches are given as [supplementary table S1, Supplementary Material](#) online. The nonsynonymous ( $K_a$ ) and synonymous ( $K_s$ ) divergences of the homologous pairs were estimated using maximum likelihood. With these divergence estimates, we first collapsed tandem duplicates in each genome. Then, for a given pair of genomes, the orthology inference procedure began by defining one-to-one pairs of homologs below a set  $K_a$  threshold as the initial orthologs ([supplementary table S1, Supplementary Material](#) online). If either of the nearest syntenic neighbors to these anchors are also homologs, then these neighboring pairs are assigned to be orthologs. This process of identifying one-to-one homologs and extending the orthology relationships to their neighboring homologs continues until no further gene pairs meeting the criteria exist.

Here, we modified the pipeline by adding a second pass in the orthology detection process. Once the initial ortholog detection was completed, we went back and examined the ambiguous cases where one gene has multiple hits in the other genome, but where only one of those homologs gives a  $K_a$  distance to the human gene below a threshold ([supplementary table S1, Supplementary Material](#) online). We then added these pairs as anchors, again running the search process until no further orthologs are found. We also modified how the algorithm handles those remaining cases where there are clusters of genes in each genome with similarity to genes in the other and where synteny cannot resolve the resulting ambiguity. We used a  $K_a$  threshold such that when one pair of genes from the two genomes is below threshold and no other pairs of genes are below two times that threshold, we defined this closest pair to be orthologs. In pairwise comparisons against human genome, the implementation of ORIS

with these changes yielded more orthologs, sometimes many more, than our earlier approach, both for distant mammal species such as platypus and for species that are more closely related to human such as the chimpanzee. Our orthology inferences are provided as a [supplementary data set, Supplementary Material](#) online, which comprises of 554,774 pairs of orthologs among these 44 genomes.

### Ancestral State Inference

We considered the presence/absence of an ortholog of each human gene in each other taxon as a character and constructed the corresponding character matrix. In this matrix, the columns are the 19,826 human genes (after the merging of tandem duplicates), and each row represents the presence or absence of an ortholog of that gene in another amniote genome (43 in total). [Supplementary figure S1, Supplementary Material](#) online, gives a schematic of our ancestral state reconstruction inferences. To make these estimates, we inferred the most recent common ancestors (MRCAs) of orthologs by placing the origin of that gene at the most recent internal node in the tree such that every species possessing the gene is a descendant of that node. This approach corresponds to making the assumption that each ortholog can appear only once in the tree, an assumption we are comfortable with due to the robustness of our synteny-based approach for orthology inference.

### Branch Length and Branch-Specific Selective Constraint Estimation

Of the 19,826 human genes analyzed, 948 had orthologs in all 43 other genomes. We obtained the longest transcripts of these 948 orthologous genes from every genome and then translated them into protein sequences. Multiple sequence alignments of the protein sequences were performed using T-coffee with default settings (Notredame et al. 2000). We then made codon-preserving nucleotide alignments from the protein alignments. After removing gaps from those alignments, we concatenated them into one meta-alignment, which had 44 taxa and 29,293 codons. A branch-specific  $\omega$  (dN/dS) was estimated by running two rounds of codeml analyses in PAML 4.9 (Yang 2007) with CodonFreq F3X4. In the first run, we estimated one universal  $\omega$  value for the entire tree and obtained the estimated number of synonymous substitutions (dS) on each branch. Then, in the second round, we set the initial branch lengths in the guide tree to be the dS estimates from the first step and allowed  $\omega$  to differ on each branch. Phylogenetic trees were visualized using the R packages ape (Paradis and Schliep 2018), phytools (Revell 2012), ggtree (Yu et al. 2017, 2018), and ggplot2 (Wickham 2016) in R 3.6.1 (R Core Team 2019).

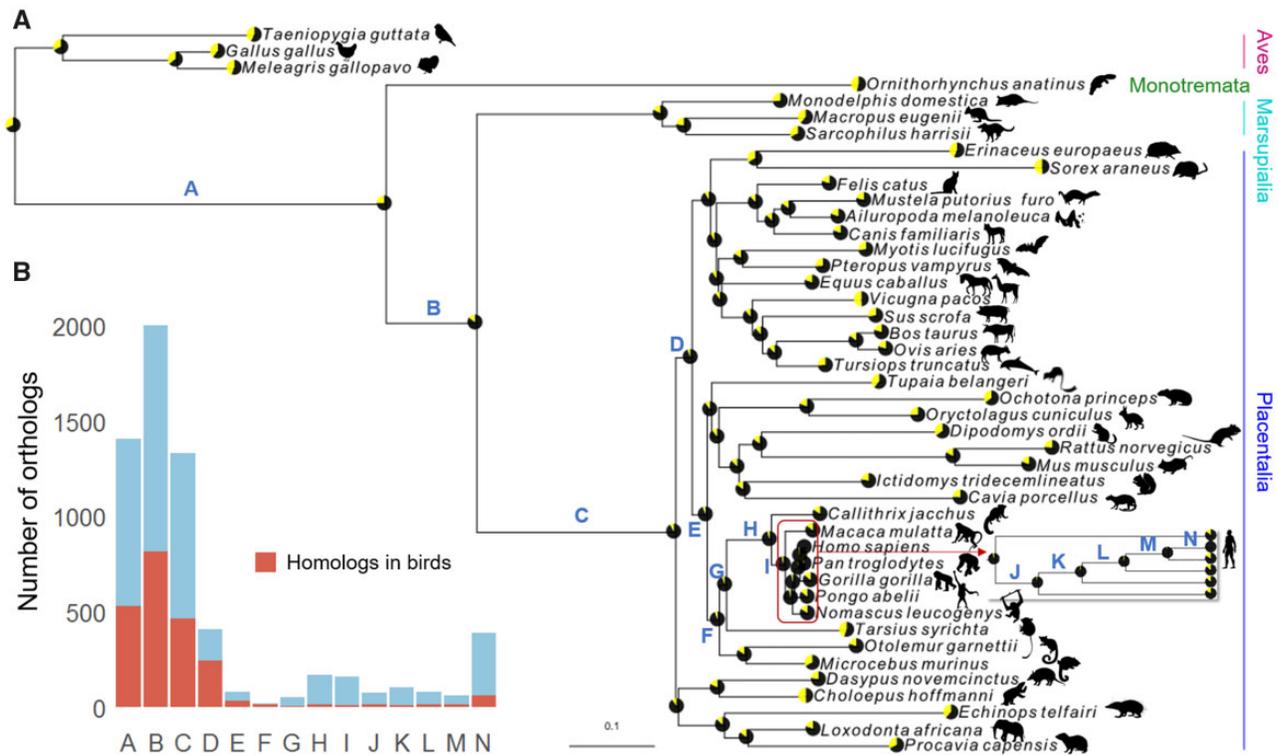
The estimated dS for each branch was converted to a time inference using the published divergence time estimate between birds and mammals of 312 Ma, as retrieved from

timetree.org (Hedges et al. 2015; Kumar et al. 2017). For branches leading to humans (A to N in fig. 1), the average rate of new orthologs appearing was estimated as the sum of number of orthologs appearing on the branch divided by the length of that branch (in millions of years). We can convert these values to an inferred rate of new ortholog occurrence on each branch, such that the number of new orthologs on each branch is given by the inferred rate multiplied by the branch length (in millions of years). We first calculated an average rate of ortholog appearance, which was given by the sum of number of orthologs on all 14 branches over the sum of their branch lengths (fig. 2a). This average rate of new ortholog appearance was 20.32 genes per million years, a value comparable with previous estimates (Long et al. 2013). Because the rate of new gene appearance is known to be variable (Long et al. 2013), we then estimated branch-specific rates, which are given by the number of orthologs appearing on a branch divided by its length (fig. 2b). We also compared our gene age inference results with those from a previous study (Zhang et al. 2010). Because of the difference in taxon sampling between the two studies, we estimated one rate for branches F, G, and H combined, one rate for I and J combined, one rate for L and M combined, and separate rates for the remainder of the individual branches, which allows for a comparison to the data of Zhang et al. (2010).

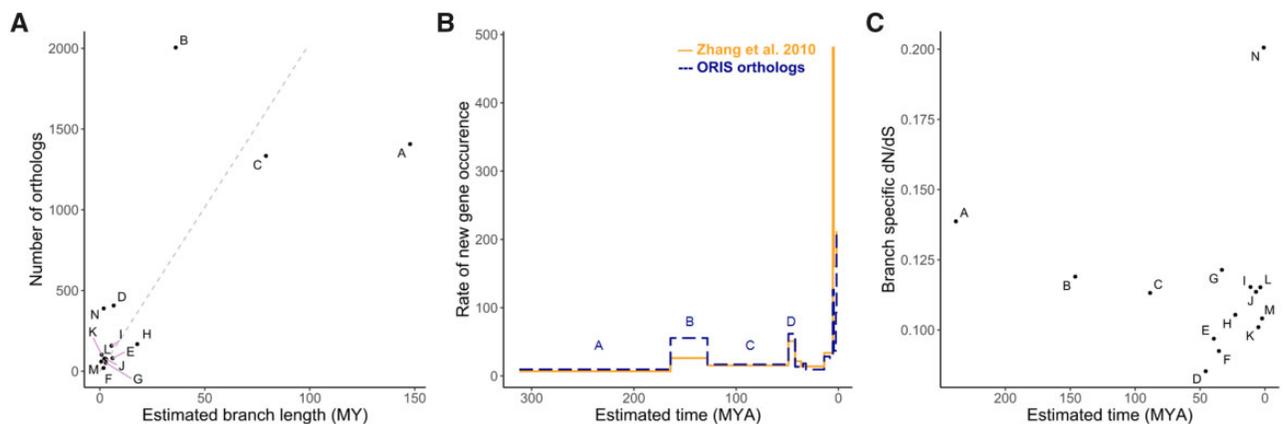
### Functional Analysis of Orthologs

Human orthologs were grouped into five bins based on their inferred first appearance on the phylogenetic tree (fig. 2). We will refer to the branch leading to the MRCA of birds and mammals as the “root branch (R),” the branch to mammalia as the “mammal branch (A),” the branch to the theria subclass (marsupials and eutherians) as the “therian branch (B)” and the branch to the stem group of eutherian mammals (i.e., the common ancestor of all placental mammals) as the “eutherian branch (C).” We merged orthologs that appeared on later branches within the eutherian mammal clade to make up a new “post-eutherian (D-N)” branch and its associated new genes. Functional annotations, including molecular functions, biological processes, pathways, protein domains, and tissue expression for each gene list, were obtained from the DAVID Bioinformatics Resources version 6.8 (Huang et al. 2009). The enrichment threshold EASE score was set to be 0.1 (Hosack et al. 2003).

Gene Ontology (GO) analyses were also performed to identify overrepresented GO terms associated with each group of orthologs using Fisher’s exact tests in PANTHER release 14.1 (Mi et al. 2019). In [supplementary table S2, Supplementary Material](#) online, we list GO terms that are overabundant among genes appearing on each internal branch of figure 1 relative to genes present on the root branch (False Discovery Rate-corrected significance level of  $P < 0.05$  [Benjamini and



**Fig. 1.**—The appearance of mammalian orthologous genes in a phylogenetic context. (a) Shown is a vertebrate orthology tree with the number of synonymous substitutions per synonymous site ( $dS$ ) as branch lengths. Black slices in the pie charts show the proportions of genes in the nonhuman genomes that have an orthologous human gene. We inferred the first appearance of each ortholog to give monophyletic groups that possess that gene (Materials and Methods): The inferred ancestral ortholog percentages are thus shown on the internal nodes.  $dS$  values were estimated using codeml with an unrooted guide tree and alignments of 948 orthologs (see Materials and Methods). The topology used was adapted from Meredith et al. (2011). The letters A–N label the internal branches leading to humans. The box region is expanded to show the primate lineage. The species images were downloaded from PhyloPic.org. (b) Bar chart of the number of orthologs on the internal branches leading to human (A–N in a). The number of such new orthologs with identified homologs (but not orthologs) in birds is shown in red (see Materials and Methods).



**Fig. 2.**—The number of the orthologs on internal branches leading to human and branch-specific selection estimates from the alignments of 948 orthologs. (a) Scatterplot of number of the orthologs versus the estimated branch length in million years (see Materials and Methods). The dashed line represents the average rate of new ortholog occurrence across the 14 branches. (b) Comparison of two estimates of the relative rate of new gene appearance on different phylogenetic branches in the mammalian tree. The blue dashed line shows the average rate of occurrence on each branch given by dividing the number of orthologs by the branch length. The yellow line is the rates estimated using new gene counts from Zhang et al. (2010) divided by the same branch lengths. (c) Branch-specific  $dN/dS$  over time: the x axis gives time (in millions of year ago) estimated from cumulative  $dS$  and the y axis gives the estimated average  $dN/dS$  for the corresponding branch (see Materials and Methods).

Hochberg 1995]). We then obtained the functional annotations for each gene list using the PANTHER Gene List Analysis option. For each branch, we counted the number of genes with molecular function GO annotations that contain any placenta-related terms, such as placental/placenta development (GO:0001890) and embryonic placenta morphogenesis (GO:0060669). Similarly, we counted the number of genes which encode “transposable-element-derived,” “retrotransposon-like,” or “transposase” proteins for each branch. The percentages of transposable-element-like genes and placenta-related genes relative to the total number of genes appearing on each branch are given in [supplementary figure S2, Supplementary Material](#) online.

We also compared the semantic similarity scores of the molecular function ontology annotations for genes from all possible pairs of our five age bins (i.e., the root, mammal, therian, eutherian, and postplacental branches). For this analysis, we used the GO annotation file from the R package `org.Hs.eg.db` (Carlson 2019). The GO semantic similarity itself was calculated using the package `GOSemSim` release 3.10 (Yu et al. 2010) in R 3.6.1, with the graph-based Wang’s method (Wang et al. 2007) and the Best-Match Average method for combining the similarity scores for GO terms in each cluster (Azuaje et al. 2005).

In order to find potential genes that are related to both tissue invasion and cancer, we uploaded each gene list to the Cancer Genome Atlas (TCGA) data portal and obtained the number of cancer census genes within each ortholog group (Cancer Genome Atlas Research Network et al. 2013; Sondka et al. 2018). Cancer hallmarks for these cancer census genes were then downloaded from the Catalogue of Somatic Mutations in Cancer (COSMIC) database release v88 (Thompson et al. 2017). From the TCGA data portal, we also downloaded, for each gene in our data set, the number of simple somatic mutations known for it. Within each age group, the number of mutations per gene was normalized by the length of the gene’s longest transcript, and the median and average mutation counts were calculated for each branch ([supplementary fig. S4, Supplementary Material](#) online). The median mutation count of each branch was compared with that of the previous branch using the Wilcoxon rank-sum test (Mann and Whitney 1947). The mutation counts were visualized using the R package `ggridges` 0.5.1 (Wilke 2018).

## Results

### Orthology Assignment and Gene Age Inference

We performed orthology inference for 43 vertebrate species’ genomes by comparison to the human genome (Materials and Methods). [Supplementary table S1, Supplementary Material](#) online, shows the parameters used for orthology inference and the number of orthologs identified in each pair of species. We used slightly relaxed homology search parameters

for orthology estimates in species more deeply diverged from humans. The pie charts at the tip nodes in [figure 1](#) show the percentage of genes with human orthologs in each genome. We next inferred the latest point of origin for every human gene in the phylogeny of [figure 1](#): in other words, the internal node of the tree that partitions all of the genomes possessing an ortholog of that gene into a monophyletic group. The number of orthologs so appearing on each internal branch is also illustrated in [figure 1](#). The percentage of genes with human orthologs under our definition is thus 60.9% in chicken (*Gallus gallus*), 54.1% in platypus (*Ornithorhynchus anatinus*), 72.9% in opossum (*Monodelphis domestica*), 80.8% in mouse (*Mus musculus*), and 85.9% in gorilla (*Gorilla gorilla*).

Next, we grouped orthologs into bins based on their phylogenetic branch of first appearance: In other words, each ortholog was inferred to have appeared on one of the internal branches leading to human. We placed 13,487 human orthologs on the root branch, the MRCA of birds and mammals, 1,407 orthologs on the mammal branch (branch A in [fig. 1](#)), 2,005 on the therian branch, 1,334 on the eutherian branch, and 1,593 on branches within the placental mammals. It is natural to ask where these genes that appeared after the split with birds came from, and in particular if they are novel duplications of genes with ancient homologs in birds. We thus examined the subgroup of human genes that have homologs in the three bird genomes but were not assigned bird orthologs by our pipeline. We computed the proportion of all genes appearing on each later branch of the tree that fell into this set of genes with bird homologs ([fig. 1b](#)). These comparisons allow us to begin to assess the relative importance of three explanations for the appearance of an orthology relationship on a particular post-bird branch: that it results from 1) a new duplication of a gene with existing homologs in birds, 2) an ancient gene whose orthologs in birds were not identified due to high sequence divergence, domain rearrangement, or annotation artifacts, or 3) a gene that lacks homologs in birds because it was produced by *de novo* processes, TE-derived open reading frames, or horizontal gene transfer (Knowles and McLysaght 2009; Syvanen 2012). Based on annotated genomes alone, we cannot distinguish explanations #2 and #3, but as [figure 1b](#) suggests, the genes appearing on the oldest post-bird branches show a relatively large fraction (~74.3%) of genes with homologs in birds, suggesting that many of these genes are duplications of existing vertebrate genes, especially because the necessity of using stringent homology cutoffs leads us to believe that we have underestimated the proportion of human genes with homologs in birds. As we move to more recent branches of the tree, relatively few genes that appear have bird homologs, which is expected, because for a gene to appear on these latter branches, its ortholog must be absent not merely in birds but also in, for instance, monotremes, marsupials, and

early diverging placental mammals such as elephants. Hence, it is reasonable to attribute these (relatively rare) newer genes to the other processes described in #3 above.

At a higher level, these results confirm that our pipeline generally meets our design goal of generating ortholog sets with very few false positives (pairs of genes identified as orthologs that actually diverged through other events such as gene duplications) at the expense of some lack of sensitivity to detecting the most diverged ortholog pairs in distantly related genomes. The inclusion of the phylogenetic ancestral reconstruction further reinforces this conclusion, because the absence of a gene from a single genome will generally be “corrected for” by the presence of other genomes equally related to humans that possess that gene. For example, the three marsupial genomes (opossum, tamar wallaby, and Tasmanian devil) lack orthologs to 27.1%, 38.6%, and 31.6% of human genes, respectively. However, they collectively lack orthologs to only 18.0% of human genes (fig. 1a).

#### The Number of Orthologs Appearing on the Eutherian Branch Is Not Unusual

The branch lengths in figure 1a are the mean number of synonymous substitutions per synonymous site (dS) across a set of 948 orthologs that are identified across all 44 species (Materials and Methods). These estimates are consistent with known species divergence time estimates (Hedges et al. 2015). We estimated branch-specific selective constraints (dN/dS or  $\omega$ ) using codeml (Yang 2007), finding that the dN/dS estimates for all of the internal branches indicate that these orthologs experience (on average) purifying selection (fig. 2).

We further asked if the therian or eutherian branches showed an unusually high number of new orthologs appearing along them, given their estimated length in terms of dS. We therefore compared the timespan of each branch (extrapolated from the dS estimates using the mammal–bird divergence time of 320 Ma as the single calibration point) with the number of orthologs appearing along that branch (fig. 2a). As shown in figure 2b, we also compared our relative estimates of branch-specific gene appearances with those from Zhang et al. (2010), controlling for the differing total numbers of genes analyzed. Notably, we do not observe an unexpectedly large number of orthologs appearing along the eutherian branch. However, both our results and those of Zhang et al. (2010) did show two trends: 1) the branches immediately before and immediately after the common placental mammal branch do show elevated rates of gene appearance and 2) a relatively large number of apparent new genes are seen along the human-specific branch, an observation that we would tend to attribute to the extensive annotation work done on the human genome rather than to any actual evolutionary trend (fig. 2). We do not see any unusual patterns of selective constraint on the common mammal branch, the therian

branch, or on the eutherian branch (fig. 2c), indicating that the evolution of mammalian pregnancy was not associated with a genome-wide “burst” of gene duplication or a relaxation of selective constraints.

#### The Majority of Human Placenta-Related Orthologs Appeared before the Origin of Placental Mammals

We named the five groups by the internal branch they were placed on, that is, orthologs on the root branch (R), mammal branch (A), therian branch (B), eutherian branch (C), and branches within the eutherian lineage (D to N). We obtained functional annotations for all human genes appearing on each branch from the PANTHER database (release 14.1).

We first compared the percentage of genes that are associated with placenta-related biological process GO terms on each branch (supplementary fig. S2, Supplementary Material online). We found that most of the placenta-related orthologs appeared on the root branch (72 genes out of 91), indicating that ancient genes were likely being repurposed to contribute to mammalian pregnancy after the origin of mammals (Stearns and Medzhitov 2016). The proportion of all genes that are annotated as placenta related is similar for the root branch and the eutherian branch (supplementary fig. S2, Supplementary Material online), again suggesting no particular process of genic innovation coincident with the evolution of placentation. Notably, only a handful of placenta-related genes were inferred to have appeared *after* the eutherian branch, reinforcing our confidence in our gene age estimates.

We also calculated the percentage of genes that encode retrotransposon-like, transposable-element-like, or transposable-element-derived proteins on each branch (supplementary fig. S2, Supplementary Material online). Interestingly, the orthologs appearing on the eutherian branch encompass most of the TE-related genes so identified. These retrotransposon-like elements could be survivors of ancient host–viral interactions that were retained in eutherian genomes. It is possible that these TE-like protein-coding genes also contributed to the origin of placental mammals through neofunctionalization (Brandt et al. 2005).

#### Orthologs Appearing on the Eutherian Branch Are Enriched for Functions in Transcriptional Regulation and Immune Responses

We next obtained functional annotations for orthologs that appeared on the internal branches of interest (A, B, C, and D–N). Our analysis used five functional annotation sets in DAVID: *UP\_Keywords*, which provides functional summaries, *GOTERM\_MF\_DIRECT* (molecular functions), *GOTERM\_BP\_DIRECT* (biological processes) for GO information, *KEGG* for biochemical pathways, and *Interpro* for protein domains (Huang et al. 2009). Functionally related annotation terms were clustered together, and an enrichment score was calculated for each cluster (Huang et al.

**Table 1**

Summary of Top Ten Functional Annotation Clusters for Each Group of Orthologs

	Top 10 Annotation Clusters	Functional Description	DAVID Enrichment Score
Orthologs on the mammalian branch	1	SPRY domain, B-box ZNF, RING-type ZNF	5.45
	2	Innate immune response	3.6
	3	Cell adhesion molecule binding, cell recognition	3.01
	4	Adaptive immune response, cytokine, interferon	2.95
	5	Interleukin, interleukin receptor binding	2.67
	6	EF-hand domain	2.52
	7	Chemical carcinogenesis, drug metabolism—cytochrome P450	2.41
	8	Steroid metabolic process, sulfotransferase activity	2.07
	9	Phototransduction guanylate cyclase activity	2.02
	10	Biom mineralization, biomineral tissue development	1.93
Orthologs on the therian branch	1	Mammalian taste receptor activity	10.15
	2	Glycoprotein, disulfide bond	9.41
	3	Olfactory and sensory transduction, GPCR activity	8.19
	4	Glycoprotein, transmembrane helix	7.1
	5	Keratinization, peptide crosslinking	4.01
	6	KRAB, C2H2-ZNF, transcription regulation	3.4
	7	Peptidase S1 activity	3.25
	8	C-type lectin, carbohydrate binding	2.56
	9	DNA binding HTH domain, endonuclease	2.11
	10	Herpes, measles, influenza A related pathway	2.09
Orthologs on the eutherian branch	1	KRAB, C2H2-ZNF, transcription regulation	28.51
	2	MAGE protein, tumor antigen	20.78
	3	Innate immune response, defense response to bacterium, $\beta$ -defensin	17.46
	4	Olfactory receptor, sensory transduction, GPCR activity	17.05
	5	Keratin, intermediate filament protein	11.36
	6	C-type lectin, carbohydrate binding	7.06
	7	Immune response, cytokine, cytokine receptor	4.6
	8	Immunoglobulin domain	4.2
	9	MHC I/II like antigen recognition protein, natural killer cell activity	4.16
	10	Keratinization, peptide crosslinking	3.66
Orthologs on the branches postplacental	1	Olfactory transduction, GPCR activity	32.51
	2	Protein deubiquitination, peptidase C19	13.21
	3	Histone, epigenetic regulation of gene expression, transcriptional misregulation in cancer, viral carcinogenesis	10.93
	4	KRAB, C2H2-ZNF, transcription regulation	8.79
	5	Innate immune response, antibacterial humoral response	7.87
	6	Defense response to bacterium, $\beta$ -defensin	7.14
	7	Cadherin, cell–cell adhesion	3.92
	8	Fungicide, defense response to fungus	3.04
	9	GRIP, protein targeting Golgi	3.03
	10	Serotonin pathway, neurotransmitter receptor activity	2.19

NOTE.—SPRY, domain in SPLa and the RYanodine receptor; RING, really interesting new gene; EF-hand, helix-loop-helix domain; GPCR, G-protein coupled receptor; KRAB, Krüppel-associated box; C2H2-ZNF, C2H2 zinc finger; HTH, helix-turn-helix; MAGE, melanoma-associated antigen; GRIP, glutamate receptor-interacting protein.

2007). The top ten enriched functional clusters for each ortholog group are shown in [table 1](#).

The most significantly enriched functional cluster for the orthologs appearing on the eutherian branch contains annotation terms such as “KRAB domain,” “DNA binding,” and “transcriptional regulation” ([table 1](#)). KRAB domains and KRAB-ZFPs are a family of transcription factors known to be controllers of sequences derived from transposable elements.

They are involved in embryonic development and genomic imprinting (Cosby et al. 2019; Shi et al. 2019). Moreover, multiple rounds of KRAB-ZFP turnovers have occurred during different phases of mammalian evolution (Huntley et al. 2006; Ecco et al. 2016; Yang et al. 2017). In fact, KRAB-ZFP family expansion has occurred independently in every vertebrate lineage (Liu et al. 2014). The enrichment of KRAB-ZFPs at the eutherian branch could therefore suggest that there were

changes in transcription regulation that co-opted sequences from TEs, including retroviruses, for the regulation of internal embryo development, as has been previously argued (Lynch et al. 2011, 2015). We also detected enrichment of the MAGE protein family, which were dramatically expanded in the mammal lineage (Ferretti et al. 2007; Weon and Potts 2015). Type I MAGE genes are often expressed in tumor cells and are involved with cancer metastasis, the cellular mechanism of which parallels placenta invasiveness (D'Souza and Wagner 2014). MAGE-like proteins are also involved in embryo implantation (Chomez et al. 2001). Orthologs on the eutherian branch were also enriched in functions including innate immune responses, natural killer cell activity, and MHC-like antigen recognition. Other research suggests that these immune-related genes might play an important role in preventing maternal–fetal immune rejection during pregnancy (Faas and de Vos 2017, 2018). Another enriched annotation cluster was related to olfactory transduction, probably due to an expansion of olfactory receptor genes during the evolution of eutherian mammals. This observation could indicate the occurrence of important physiological functional changes but is plausibly unrelated to the evolution of the placenta (Niimura et al. 2014).

Tests of enrichment of protein families comparing the ancient orthologs (those appearing on the root branch) to orthologs appearing on the mammal branch, therian branch, and eutherian branch yielded similar results (supplementary table S2, Supplementary Material online). KRAB transcription factors were enriched upon the appearance of placentation on the eutherian branch. Overrepresentation of these zinc-finger transcription factors could indicate that KRAB-mediated rewiring of genetic regulatory networks played an important role during the evolution of mammalian pregnancy. Cell adhesion molecules and immunoglobulin proteins were enriched in the mammal lineage, suggesting changes in immune modulation to avoid maternal–fetal immune conflict even at the early stages of mammalian evolution. Curiously, when we looked quite broadly at the functional roles of these genes of differing ages using the semantic similarity of the molecular function GO terms, we found that orthologs appearing on the eutherian branch have the most distinct GO terms relative to the other four groups (supplementary fig. S3, Supplementary Material online).

### Gene Age as a Predictor of Tissue Invasion

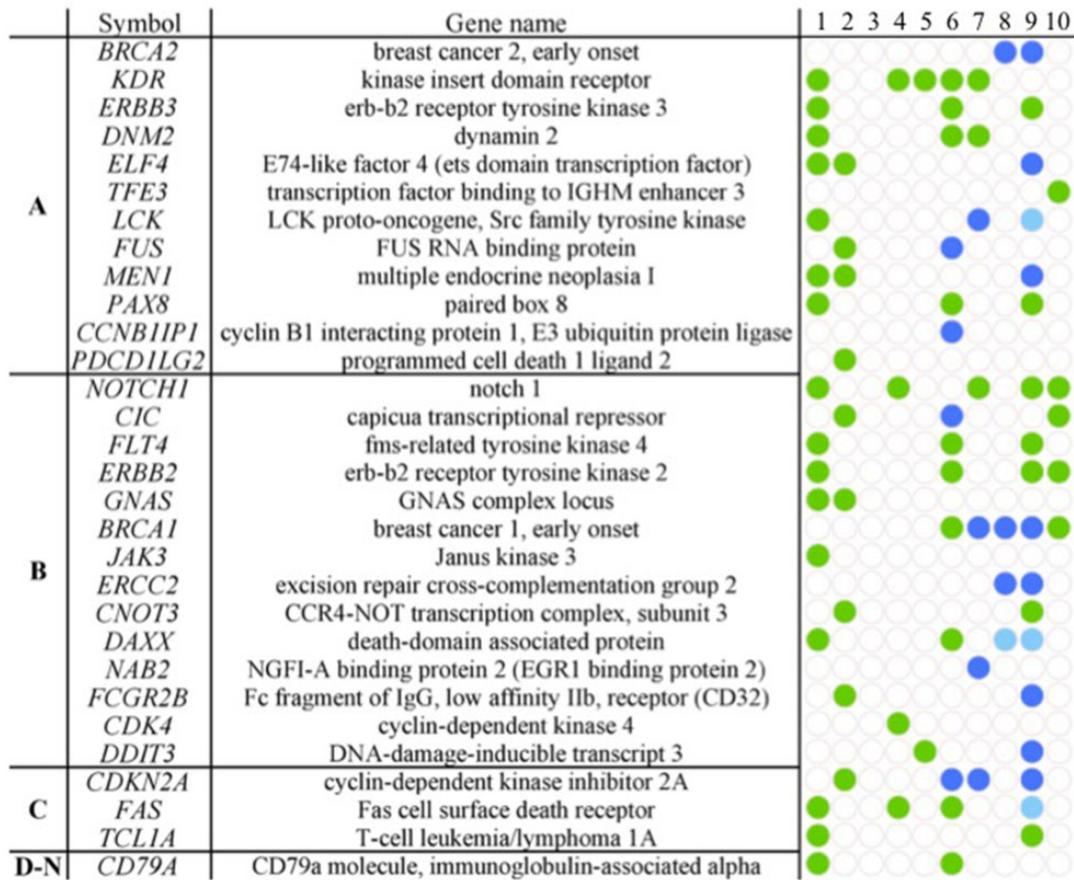
In human and other eutherian mammals with hemochorial placenta, the trophoblasts will invade the uterine lining with behavior similar to metastatic cancer cells (e.g., proliferative signaling, evasion of apoptosis, and tissue invasion; Holtan et al. 2009). Figure 3 shows orthologs from each internal branch that the cancer census has annotated with cancer hallmarks from the COSMIC database (Hanahan and Weinberg 2011; Thompson et al. 2017). Cancer census genes

with unknown hallmarks are listed in supplementary table S3, Supplementary Material online. The eutherian branch is again not unusual in the prevalence of cancer-associated genes that appeared along it. Instead, there is a direct relationship between the age of an ortholog and its tendency to be associated with cancer, with older orthologs showing the higher prevalence (table 2, Fisher's exact tests). However, we observed a trend of an increasing relative prevalence of somatic mutations when moving from the root branch to the eutherian one (supplementary fig. S4, Supplementary Material online), where this average somatic mutation propensity peaked. In other words, the orthologs appearing around the time of the eutherian origin are more often observed to have somatic mutations in humans, though we lack a clear hypothesis as to why. These results also complement previous findings that mutations in the very ancient genes shared by all metazoans can disrupt necessary interactions between those old genes and younger genes, driving inherited diseases or cancer (Domazet-Lošo and Tautz 2008; Trigos et al. 2019).

At the level of specific genes, we found multiple cancer census genes that appeared on the therian and eutherian branches that are most likely involved with placenta formation. For example, on the therian branch, *NOTCH1* is involved in the Notch signaling pathway, which is crucial for maternal–fetal communication (Cuman et al. 2014); the *ERBB2* (erb-b2 receptor tyrosine kinase) gene is a member of the epidermal growth factor receptor family, which was previously shown to be expressed in both normal trophoblasts and malignant breast/ovary tumors (Ferretti et al. 2007). On the eutherian branch, the *FAS* gene is involved in apoptotic processes, and it has been found under significant positive selection in mammalian evolution (Vicens and Posada 2018). These findings hint at shared circuits between placenta invasion and cancer metastasis.

### Discussion

We illustrate the power of high-resolution comparative genomics to illuminate functional patterns in mammalian genomes. The evolutionary origin of mammalian pregnancy was a complex process involving maternal–fetal tissue fusion, gene regulatory network rewiring, as well as the emergence of a new cell type (decidual stromal cells) to repress the immune conflict between mother and developing offspring (Wildman et al. 2006; Lynch et al. 2015; Chavan et al. 2016). The predictions from our orthology inferences tend to support previous findings and illustrate again the value of the comparative method in resolving the origins of complex traits (Wildman 2016). Most importantly, it appears that genes that allow for the key molecular and cellular events in placentation were preexisting and co-opted in the evolution of placenta (Knox and Baker 2008; Chavan et al. 2016; Guernsey et al. 2017). A novel observation of our analyses is that although these “structural” genes were likely



**Fig. 3.**—Cancer hallmark genes for orthologs appearing on the mammal, therian, eutherian, and postplacental branches of the phylogeny in figure 1. Four groups of genes are listed based on their evolutionary time of appearance. Branch A is the mammal branch, B is the therian branch, C is the eutherian branch, and D–N are the postplacental branches. The association of these genes with cancer hallmarks were shown. Green circles stand for “promotes,” dark blue circles represent “suppresses,” and the aqua circles stand for both. The cancer hallmark annotations for cancer census genes were obtained from COSMIC release v88 (Hanahan and Weinberg 2011; Thompson et al. 2017). Ten cancer hallmarks are 1, proliferative signaling; 2, suppression of growth; 3, escaping immune response to cancer; 4, cell replicative immortality; 5, tumor promoting inflammation; 6, invasion and metastasis; 7, angiogenesis; 8, genome instability and mutations; 9, escaping programmed cell death; 10, change of cellular energetics.

**Table 2**

Number of Orthologs That Are Expressed in Placenta (from DAVID) and Are Annotated as Census Cancer Genes in COSMIC

Branch	Number of Genes	Number of Genes Expressed in Placenta <sup>a</sup>	Number of Cancer Census Genes <sup>a</sup>
Root branch	13,472	7,215**	501**
Mammal branch	1,403	616**	24*
Therian branch	1,996	785**	34**
Eutherian branch	1,320	388**	9
Postplacental branches	1,174	277	4

<sup>a</sup>The gene count on each branch was compared with the sum of gene counts on later branches, for example, number of orthologs on eutherian branch was compared with the number of orthologs on all branches after the eutherian branch. Fisher’s exact tests were performed to examine whether the true odds ratios are >1.

\**P* < 0.05.

\*\**P* < 0.0001.

preexisting, there was a radiation of regulatory genes coincident with the placental radiation, potentially allowing for the evolution of new expression patterns among those existing genes. This hypothesis would in some respects parallel the

observation that transposon-related sequences were repurposed at this same time to help regulate placental development (Lynch et al. 2011; Emera and Wagner 2012). Indeed, the eutherian branch shows the strongest pattern of the

development of new genes from TE substrates, suggesting the importance of this TE repurposing in placental mammal evolution. As an aside, we note that all of these estimates have resolution no finer than the total length of the common eutherian branch: Our results do not argue for an unusually rapid origin of such regulatory innovations (fig. 1).

This use of preexisting genes for new purposes is a very common, if perhaps insufficiently recognized, feature of evolution. Morphological examples include not only Gould's famous panda's thumb but also the skin glands responsible for milk production (Gould 1980; Oftedal 2012). At a molecular level, another part of the mammalian trait repertoire also shows strong evidence of such repurposing: a number of proteins seem to have been co-opted during the evolution of milk fat globule production, in one case perhaps through the employment of an enzyme in a nonenzymatic role (Vorbach et al. 2002; Ogg et al. 2004; Oftedal 2012). Although none of the mammalian innovations discussed involve the duplication of the genes, there are clear parallels to that process: In both cases, the evolutionary innovation is (not surprisingly) based on the co-option of either primary or secondary protein activities (Conant and Wolfe 2008).

We have also explored the power of our comparative genomic pipeline for uncovering new features of the genomics of cancer by inferring a list of genes that are likely involved in tissue invasion in both placenta and tumors (Costanzo et al. 2018). As seen in the analysis of placental genes, the eutherian branch is not unusual in the proportion of cancer-associated genes originating along it: Instead the genes so involved are generally evolutionarily ancient. In terms of somatic mutation counts, genes that appeared around the eutherian origin show slightly elevated average number of mutations, perhaps representing a tradeoff between the intolerance of older genes to such mutations and the chances of a mutation having a cell-lineage level selective effect. In this study, we only reported well-annotated cancer census genes originating on each branch. However, this criterion might be too stringent, with more genes related to cancer metastasis and placenta formation yet to be explored. Nonetheless, we did find some known associations: our pipeline finds that the placenta-specific protein 1 (*PLAC1*) gene arose at the origin of eutherian mammals, which agrees with previous findings (Devor et al. 2014). Notably, *PLAC1* is a promising target for development of antibody-drug conjugates (Nejadmoghaddam et al. 2017; Yuan et al. 2018). A more striking finding is that even within the confined temporal window of the mammalian radiation, phylogenetically older genes are more likely to show associations with cancer.

More generally, as comparative genomics resources continue to improve, researchers in many areas of biological research will find it fruitful to refine their hypotheses regarding gene function in light of the evolutionary history of the genes in question: It makes little sense to explore old and

evolutionarily conserved pathways by studying species-specific genes. On the other hand, the work above reminds us of the opportunistic and unexpected manner in which evolution can repurpose old genes to new functions. Detecting such innovations remains an important open problem in molecular evolution.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

We would like to thank V. Lynch, J.C. Pires, M. Lucy, T. Spencer, and J. Thorne for helpful conversations and H. Truong for computational advice. This work was supported by National Science Foundation Grants NSF-IOS-1339156 (Y.H. and G.C.C.), NSF-DBI-1358997 (M.B., H.J.L., K.B., T.M., and G.C.C.), NSF-CCF-1421765 (G.C.C.), European Union's Seventh Framework Programme Grant 613689 (Y.H. and G.C.C.), and by the Triangle Comparative and Evolutionary Medicine Center (Y.H.).

## Literature Cited

- Azuaje F, Wang H, Bodenreider O. 2005. Ontology-driven similarity approaches to supporting gene functional assessment. In: Lord P, Stevens R, Butler JA, McEntire R, editors. Proceedings of the ISMB'2005 SIG Meet Bio-Ontologies; Detroit. p. 9–10.
- Bekaert M, Conant GC. 2011. Copy number alterations among mammalian enzymes cluster in the metabolic network. *Mol Biol Evol.* 28:1111–1121.
- Bekaert M, Conant GC. 2014. Gene duplication and phenotypic changes in the evolution of mammalian metabolic networks. *PLoS One* 9:e87115.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc.* 57:289–300.
- Brandt J, Veith AM, Volf JN. 2005. A family of neofunctionalized Ty3/gypsy retrotransposon genes in mammalian genomes. *Cytogenet Genome Res.* 110:307–317.
- Cancer Genome Atlas Research Network, et al. 2013. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 45(10):1113–1120.
- Carlson M. 2019. org.Hs.eg.db: genome wide annotation for human. R package version 3.8.2.
- Chavan AR, Bhullar BAS, Wagner GP. 2016. What was the ancestral function of decidual stromal cells? A model for the evolution of eutherian pregnancy. *Placenta* 40:40–51.
- Chomez P, et al. 2001. An overview of the MAGE gene family with the identification of all human members of the family. *Cancer Res.* 61:5544–5551.
- Churakov G, et al. 2009. Mosaic retroposon insertion patterns in placental mammals. *Genome Res.* 19:868–875.
- Conant GC. 2009. Neutral evolution on mammalian protein surfaces. *Trends Genet.* 25:377–381.
- Conant GC, Wagner A. 2002. GenomeHistory: a software tool and its application to fully sequenced genomes. *Nucleic Acids Res.* 30:3378–3386.

- Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet.* 9:938–950.
- Cosby RL, Chang N-C, Feschotte C. 2019. Host–transposon interactions: conflict, cooperation, and cooption. *Genes Dev.* 33:1098–1116.
- Costanzo V, Bardelli A, Siena S, Abrignani S. 2018. Exploring the links between cancer and placenta development. *Open Biol.* 8:180081.
- Cuman C, et al. 2014. Fetal-maternal communication: the role of Notch signalling in embryo implantation. *Reproduction* 147:R75–R86.
- Devor EJ, et al. 2014. Placenta-specific protein 1: a potential key to many oncofetal-placental OB/GYN research questions. *Obstet Gynecol Int.* 2014:678984.
- Domazet-Lošo T, Tautz D. 2008. An ancient evolutionary origin of genes associated with human genetic diseases. *Mol Biol Evol.* 25:2699–2707.
- dos Reis M, et al. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc Biol Sci.* 279:3491–3500.
- D’Souza AW, Wagner GP. 2014. Malignant cancer and invasive placentation: a case for positive pleiotropy between endometrial and malignancy phenotypes. *Evol Med Public Heal.* 2014:136–145.
- Dunwell TL, Paps J, Holland PWH. 2017. Novel and divergent genes in the evolution of placental mammals. *Proc Biol Sci.* 284:20171357.
- Ecco G, et al. 2016. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. *Dev Cell* 36:611–623.
- Emera D, Wagner GP. 2012. Transposable element recruitments in the mammalian placenta: impacts and mechanisms. *Brief Funct Genomics.* 11:267–276.
- Faas MM, de Vos P. 2017. Uterine NK cells and macrophages in pregnancy. *Placenta* 56:44–52.
- Faas MM, de Vos P. 2018. Innate immune cells in the placental bed in healthy pregnancy and preeclampsia. *Placenta* 69:125–133.
- Ferretti C, Bruni L, Dangles-Marie V, Pecking AP, Bellet D. 2007. Molecular circuits shared by placental and cancer cells, and their implications in the proliferative, invasive and migratory capacities of trophoblasts. *Hum Reprod Update* 13:121–141.
- Gould SJ. 1980. *The panda’s thumb: more reflections in natural history.* New York: W.W. Norton & Company.
- Guernsey MW, Chuong EB, Cornelis G, Renfree MB, Baker JC. 2017. Molecular conservation of marsupial and eutherian placentation and lactation. *Elife* 6:e27450.
- Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* 144:646–674.
- Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. 2015. Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol.* 32:835–845.
- Holtan SG, Creedon DJ, Haluska P, Markovic SN. 2009. Cancer and pregnancy: parallels in growth, invasion, and immune modulation and implications for cancer therapeutic agents. *Mayo Clin Proc.* 84:985–1000.
- Hosack DA, Dennis G, Sherman BT, Lane HC, Lempicki RA. 2003. Identifying biological themes within lists of genes with EASE. *Genome Biol.* 4:R70.
- Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4:44–57.
- Huang DW, et al. 2007. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* 8:R183.
- Huntley S, et al. 2006. A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res.* 16:669–677.
- Kim M, et al. 2016. GENEStation 1.0: a synthetic resource of diverse evolutionary and functional genomic data for studying the evolution of pregnancy-associated tissues and phenotypes. *Nucleic Acids Res.* 44:D908–D916.
- Knowles DG, McLysaght A. 2009. Recent de novo origin of human protein-coding genes. *Genome Res.* 19:1752–1759.
- Knox K, Baker JC. 2008. Genomic evolution of the placenta using cooption and duplication and divergence. *Genome Res.* 18:695–705.
- Koonin EV. 2005. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet.* 39:309–338.
- Koonin EV, Galperin M. 2003. *Sequence—evolution—function: computational approaches in comparative genomics.* Boston (MA): Kluwer Academic.
- Kriegs JO, et al. 2006. Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol.* 4:e91.
- Kristensen DM, Wolf YI, Mushegian AR, Koonin EV. 2011. Computational methods for Gene Orthology inference. *Brief Bioinform.* 12:379–391.
- Kriventseva EV, Rahman N, Espinosa O, Zdobnov EM. 2008. OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nucleic Acids Res.* 36:D271–D275.
- Kriventseva EV, et al. 2015. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res.* 43:D250–D256.
- Kshitiz, et al. 2019. Evolution of placental invasion and cancer metastasis are causally linked. *Nat Ecol Evol.* 3:1743–1753.
- Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol.* 34:1812–1819.
- Lee A, Nolan A, Watson J, Tristem M. 2013. Identification of an ancient endogenous retrovirus, predating the divergence of the placental mammals. *Philos Trans R Soc B Biol Sci.* 368:20120503.
- Lefèvre CM, Sharp JA, Nicholas KR. 2010. Evolution of lactation: ancient origin and extreme adaptations of the lactation system. *Annu Rev Genom Hum Genet.* 11:219–238.
- Liu H, Chang LH, Sun Y, Lu X, Stubbs L. 2014. Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. *Genome Biol Evol.* 6:510–525.
- Long M, VanKuren NW, Chen S, Vrbancovski MD. 2013. New gene evolution: little did we know. *Annu Rev Genet.* 47:307–333.
- Lynch VJ, Leclerc RD, May G, Wagner GP. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet.* 43:1154–1159.
- Lynch VJ, et al. 2015. Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. *Cell Rep.* 10:551–561.
- Mann HB, Whitney DR. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Statist.* 18:50–60.
- Meredith RW, et al. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* 334:521–524.
- Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47:D419–D426.
- Mincheva-Nilsson L, Baranov V. 2014. Placenta-derived exosomes and syncytiotrophoblast microparticles and their role in human reproduction: immune modulation for pregnancy success. *Am J Reprod Immunol.* 72:440–457.
- Mossman HW. 1937. Comparative morphogenesis of the fetal membranes and accessory uterine structures. *Contrib Embryol.* 26:129–246.
- Mossman HW. 1987. *Vertebrate fetal membranes: comparative ontogeny and morphology; evolution; phylogenetic significance; basic functions; research opportunities.* New Brunswick (Canada): Rutgers University Press.
- Nejadmoghadam MR, et al. 2017. Placenta-specific1 (PLAC1) is a potential target for antibody–drug conjugate-based prostate cancer immunotherapy. *Sci Rep.* 7:13373.
- Niimura Y, Matsui A, Touhara K. 2014. Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary

- dynamics of orthologous gene groups in 13 placental mammals. *Genome Res.* 24:1485–1496.
- Notredame C, Higgins DG, Heringa J. 2000. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 302(1):205–217.
- Oftedal OT. 2012. The evolution of milk secretion and its ancient origins. *Animal* 6(3):355–368.
- Ogg SL, Dobbie L, Weldon AK, Mather IH, Smith A. 2004. Expression of butyrophilin (Bt1a1) in lactating mammary gland is essential for the regulated secretion of milk-lipid droplets. *Proc Natl Acad Sci U S A.* 101(27):10084–10089.
- O’Leary MA, et al. 2013. The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* 339:662–667.
- Paradis E, Schliep K. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2018:1–3.
- R Core Team. 2019. R: A language and environment for statistical computing.
- Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* 3(2):217–223.
- Roberts RM, Green JA, Schulz LC. 2016. The evolution of the placenta. *Reproduction* 152(5):R179–R189.
- Shi H, et al. 2019. ZFP57 regulation of transposable elements and gene expression within and beyond imprinted domains. *Epigenet Chromatin.* 12(1):49.
- Smith KK. 2015. Placental evolution in therian mammals. In: *Great transformations in vertebrate evolution*. Chicago (IL): The University of Chicago Press. p. 205–225.
- Sondka Z, et al. 2018. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* 18(11):696–705.
- Stearns SC, Medzhitov R. 2016. *Evolutionary medicine*. Sunderland (MA): Sinauer Associates, Inc.
- Syvanen M. 2012. Evolutionary implications of horizontal gene transfer. *Annu Rev Genet.* 46(1):341–358.
- Tarver JE, et al. 2016. The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome Biol Evol.* 8(2):330–344.
- Thompson S, et al. 2017. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 45:D777–D783.
- Trigos AS, Pearson RB, Papenfuss AT, Goode DL. 2019. Somatic mutations in early metazoan genes disrupt regulatory links between unicellular and multicellular genes in cancer. *Elife* 8:e40947.
- Van Dyke JU, Brandley MC, Thompson MB. 2014. The evolution of viviparity: molecular and genomic data from squamate reptiles advance understanding of live birth in amniotes. *Reproduction* 147:R15–R26.
- Vicens A, Posada D. 2018. Selective pressures on human cancer genes along the evolution of mammals. *Genes* 9:582.
- Vogel P. 2005. The current molecular phylogeny of Eutherian mammals challenges previous interpretations of placental evolution. *Placenta* 26:591–596.
- Vorbach C, Scriven A, Capecchi MR. 2002. The housekeeping gene xanthine oxidoreductase is necessary for milk fat droplet enveloping and secretion: gene sharing in the lactating mammary gland. *Genes Dev.* 16(24):3223–3235.
- Wagner GP, Kin K, Muglia L, Pavličev M. 2014. Evolution of mammalian pregnancy and the origin of the decidual stromal cell. *Int J Dev Biol.* 58:117–126.
- Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF. 2007. A new method to measure the semantic similarity of GO terms. *Bioinformatics* 23(10):1274–1281.
- Ward N, Moreno-Hagelsieb G. 2014. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss? *PLoS One* 9(7):e101850.
- Weon JL, Potts PR. 2015. The MAGE protein family and cancer. *Curr Opin Cell Biol.* 37:1–8.
- Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. 2nd ed. New York: Springer-Verlag.
- Wildman DE. 2016. IFA award in placentology lecture: phylogenomic origins and evolution of the mammalian placenta. *Placenta* 48:S31–S39.
- Wildman DE, et al. 2006. Evolution of the mammalian placenta revealed by phylogenetic analysis. *Proc Natl Acad Sci U S A.* 103(9):3203–3208.
- Wilke CO. 2018. ggrridges: ridgeline plots in ‘ggplot2’. R package version 0.5.1.
- Wu J, Yonezawa T, Kishino H. 2017. Rates of molecular evolution suggest natural history of life history traits and a post-K-Pg nocturnal bottleneck of placentals. *Curr Biol.* 27(19):3025–3033.
- Yang P, Wang Y, Macfarlan TS. 2017. The role of KRAB-ZFPs in transposable element repression and mammalian evolution. *Trends Genet.* 33(11):871–881.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Yu G, Lam TT, Zhu H, Guan Y. 2018. Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Mol Biol Evol.* 35(12):3041–3043.
- Yu G, Smith DK, Zhu H, Guan Y, Lam T. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol.* 8(1):28–36.
- Yu G, et al. 2010. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 26(7):976–978.
- Yuan H, et al. 2018. Plac1 is a key regulator of the inflammatory response and immune tolerance in mammary tumorigenesis. *Sci Rep.* 8(1):5717.
- Zerbino DR, et al. 2018. Ensembl 2018. *Nucleic Acids Res.* 46(D1):D754–D761.
- Zhang YE, Vrbčanovski MD, Landback P, Marais GAB, Long M. 2010. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol.* 8(10):e1000494.

Associate editor: Federico Hoffmann