ARTICLE

https://doi.org/10.1038/s41586-019-1338-5

Gene expression across mammalian organ development

Margarida Cardoso-Moreira[1,2]*, Jean Halbert[2], Delphine Valloton[2], Britta Velten[3], Chunyan Chen[4,5,6], Yi Shao[4,5,6], Angélica Liechti[2], Kelly Ascenção[2], Coralie Rummel[2], Svetlana Ovchinnikova[1], Pavel V. Mazin[7,8,9], Ioannis Xenarios[2], Keith Harshman[2], Matthew Mort[10], David N. Cooper[10], Carmen Sandi[11], Michael J. Soares[12,13], Paula G. Ferreira[14,15], Sandra Afonso[16], Miguel Carneiro[16,17], James M. A. Turner[18], John L. VandeBerg[19,20], Amir Fallahshahroudi[21], Per Jensen[21], Rüdiger Behr[22,23], Steven Lisgo[24], Susan Lindsay[24], Philipp Khaitovich[7,25,26], Wolfgang Huber[3], Julie Baker[27], Simon Anders[1], Yong E. Zhang[4,5,25] & Henrik Kaessmann[1]*

The evolution of gene expression in mammalian organ development remains largely uncharacterized. Here we report the transcriptomes of seven organs (cerebrum, cerebellum, heart, kidney, liver, ovary and testis) across developmental time points from early organogenesis to adulthood for human, rhesus macaque, mouse, rat, rabbit, opossum and chicken. Comparisons of gene expression patterns identified correspondences of developmental stages across species, and differences in the timing of key events during the development of the gonads. We found that the breadth of gene expression and the extent of purifying selection gradually decrease during development, whereas the amount of positive selection and expression of new genes increase. We identified differences in the temporal trajectories of expression of individual genes across species, with brain tissues showing the smallest percentage of trajectory changes, and the liver and testis showing the largest. Our work provides a resource of developmental transcriptomes of seven organs across seven species, and comparative analyses that characterize the development and evolution of mammalian organs.

Understanding the molecular evolution of mammalian phenotypic traits is a fundamental biological goal. To achieve it, evolutionary studies need to be conducted within a developmental framework, both because adult phenotypes are defined during development[1–3] and because evolutionary and developmental processes are intertwined. von Baer noted in the nineteenth century that morphological differences between species increase as development advances[4] and supporting evidence has accumulated[4,5]. Understanding the molecular foundations of these patterns will facilitate the identification of general principles that underlie phenotypic evolution.

Here we provide a resource of bulk transcriptomes across developmental stages, covering multiple organs from early organogenesis to adulthood in seven species, enabling direct comparisons of expression patterns in organ development within and across mammals (http://evodevoapp.kaessmannlab.org). This resource enabled us to analyse the evolution of gene expression within mammalian organs across developmental stages.

Organ developmental transcriptomes

We generated gene expression time series for six mammals (human, rhesus macaque, mouse, rat, rabbit and opossum) and a bird (red junglefowl, henceforth referred to as 'chicken') using RNA sequencing. We sampled seven organs that represent the three germ layers: brain (forebrain/cerebrum) and cerebellum (hindbrain/cerebellum) (ectoderm); heart, kidney, ovary and testis (mesoderm); and liver (endoderm) (Fig. 1a). The time series span from early organogenesis to adulthood, plus senescence in primates. We sampled prenatal development at regular intervals (for example, daily in rodents, weekly in humans), and postnatally we sampled neonates, 'infants', juveniles, and adults (Fig. 1). There are exceptions: we lack early prenatal data for rhesus macaque and chicken, and ovary data for rhesus macaque and postnatal human development (Supplementary Table 1). Because organ development in marsupials occurs predominantly postnatally[6], all sampled stages except for one were collected postnatally. The dataset consists of 1,893 libraries (Supplementary Table 2).

The global relationships among all samples can be explored through a principle component analysis (PCA) (Fig. 1b). The first principal component (PC1), explaining most variance in gene expression, separates the samples by the germ layer from which organs originate. PC2 separates the samples by developmental stage (from early to late development). PC3 and PC4 separate the samples by species (Extended Data

---

[1]Center for Molecular Biology of Heidelberg University (ZMBH), Heidelberg, Germany. [2]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland. [3]Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [4]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [5]State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [6]University of Chinese Academy of Sciences, Beijing, China. [7]Center for Neurobiology and Brain Restoration, Skolkovo Institute of Science and Technology, Moscow, Russia. [8]Institute for Information Transmission Problems (Kharkevich Institute) RAS, Moscow, Russia. [9]Faculty of Computer Science, HSE University, Moscow, Russia. [10]Institute of Medical Genetics, Cardiff University, Cardiff, UK. [11]Laboratory of Behavioral Genetics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [12]Institute for Reproduction and Perinatal Research, Departments of Pathology and Laboratory Medicine and Pediatrics, University of Kansas Medical Center, Kansas City, MO, USA. [13]Center for Perinatal Research, Children's Research Institute, Children's Mercy, Kansas City, MO, USA. [14]Departamento de Anatomia, Universidade do Porto, Porto, Portugal. [15]ICBAS (Instituto de Ciências Biomédicas Abel Salazar), UMIB (Unidade Multidisciplinar de Investigação Biomédica), Universidade do Porto, Porto, Portugal. [16]CIBIO/InBIO, Centro de Investigacão em Biodiversidade e Recursos Genéticos, Universidade do Porto, Porto, Portugal. [17]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal. [18]Sex Chromosome Biology Laboratory, The Francis Crick Institute, London, UK. [19]South Texas Diabetes and Obesity Institute, School of Medicine, The University of Texas Rio Grande Valley, Brownsville, Harlingen and Edinburg, TX, USA. [20]The Department of Human Genetics, School of Medicine, The University of Texas Rio Grande Valley, Brownsville, Harlingen and Edinburg, TX, USA. [21]AVIAN Behavioural Genomics and Physiology Group, IFM Biology, Linköping University, Linköping, Sweden. [22]Platform Degenerative Diseases, German Primate Center, Leibniz Institute for Primate Research (DPZ), Göttingen, Germany. [23]DZHK (German Center for Cardiovascular Research), Partner Site Göttingen, Göttingen, Germany. [24]Human Developmental Biology Resource, Institute of Genetic Medicine, Newcastle University, Newcastle, UK. [25]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China. [26]CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [27]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. *e-mail: m.moreira@zmbh.uni-heidelberg.de; h.kaessmann@zmbh.uni-heidelberg.de

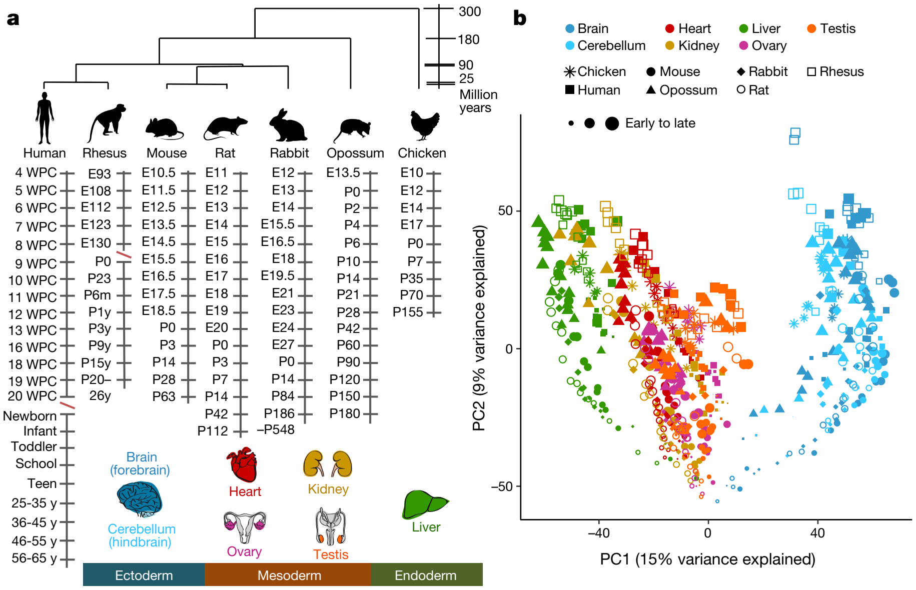


Fig. 1 | Organ developmental transcriptomes. a, Species, organs and stages sampled. Red slashes highlight two sampling gaps: human development is not covered between 20 and 38 weeks post-conception (WPC), and rhesus macaque development between embryonic (E) days 130 and 160. **b**, PCA based on 7,696 1:1 orthologues across all species. Each dot represents the median across replicates (approximately 2–4).

Fig. 1a). In PCAs of individual organs, PC1 and PC2 order the samples by developmental stage and separate them by species (Extended Data Fig. 1b). In the global PCA (Fig. 1b), the earliest samples from different organs cluster together, suggesting strong commonalities. We hypothesized that developmental programs are still largely shared across organs at these early stages and found that organ transcriptomes are indeed most similar at these stages (Extended Data Fig. 1c). Throughout development, the expression of transcription factors differs more between organs than that of the whole transcriptome (Extended Data Fig. 1c), consistent with transcription factors directing organogenesis.

Next, we identified genes with significant temporal expression changes in each organ, termed developmentally dynamic genes (DDGs; Extended Data Fig. 2a; Supplementary Tables 3–9; Methods). DDGs reflect changes during development in gene regulation, cell type abundance, and/or the proportion of cells undergoing division[1]. Consistently, between 79% and 91% of protein-coding genes in each species are DDGs (Extended Data Fig. 2b), including genes with housekeeping functions. DDGs are enriched with phenotypes associated with the development, anatomy and physiology of each organ, plus general growth and patterning (false discovery rate (FDR) < 0.01, hypergeometric test; Supplementary Tables 10, 11). DDGs are under stronger functional constraints[7–10] than non-DDGs, and the constraints increase with the number of organs in which genes show temporal dynamic expression (Extended Data Fig. 2c). The increased constraints extend to dosage changes, with DDGs being less tolerant to duplication and deletion variants[11] (Extended Data Fig. 2d).

In each species, 6–15% of the genes (Extended Data Fig. 2b) are DDGs in only one organ, and are consistently enriched with organ-specific phenotypes (FDR < 0.01, hypergeometric test; Supplementary Table 12). The fraction of expressed organ-specific DDGs increases during development (Extended Data Fig. 2e), correlating with organ differentiation and maturation. The opposite is observed for transcription factors, the contribution of which is highest earlier in development (Extended Data Fig. 2f).

Developmental correspondences and heterochrony

Embryonic development is divided into 23 Carnegie stages, which are matched across species[12–15] (Extended Data Fig. 3a). However, cross-species correspondences during fetal and postnatal development are unknown. Therefore, we used the developmental transcriptomes to establish stage correspondences across species throughout the entire development (Methods; Fig. 2a; Extended

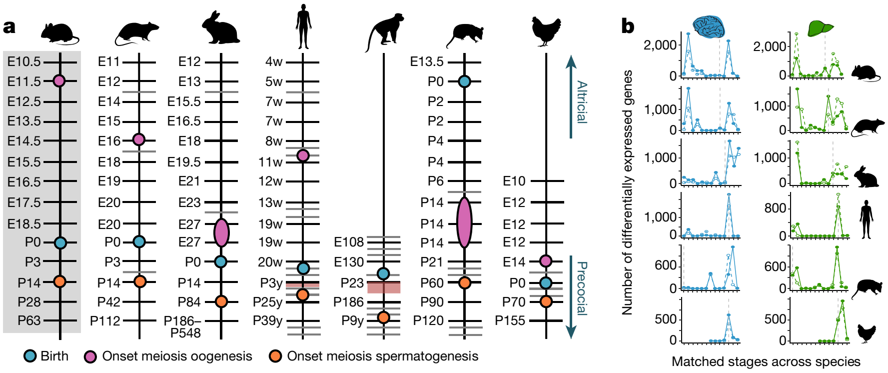


Fig. 2 | Developmental correspondences. a, Stage correspondences across species. Grey bars represent additionally sampled stages. Red shading highlights sampling gaps. In rhesus macaque, male meiosis starts at 3–4 years[36]. Because we did not detect expression of meiotic genes in the 3-year-olds, we placed the onset of meiosis between 3 and 9 years. **b**, Number of genes differentially expressed between adjacent, species-matched, stages for brain and liver ($\log_2$ fold change ≥ 0.5; other organs in Extended Data Fig. 5). Solid lines mark genes that increase in expression and dashed lines genes that decrease. Vertical dotted line marks birth.

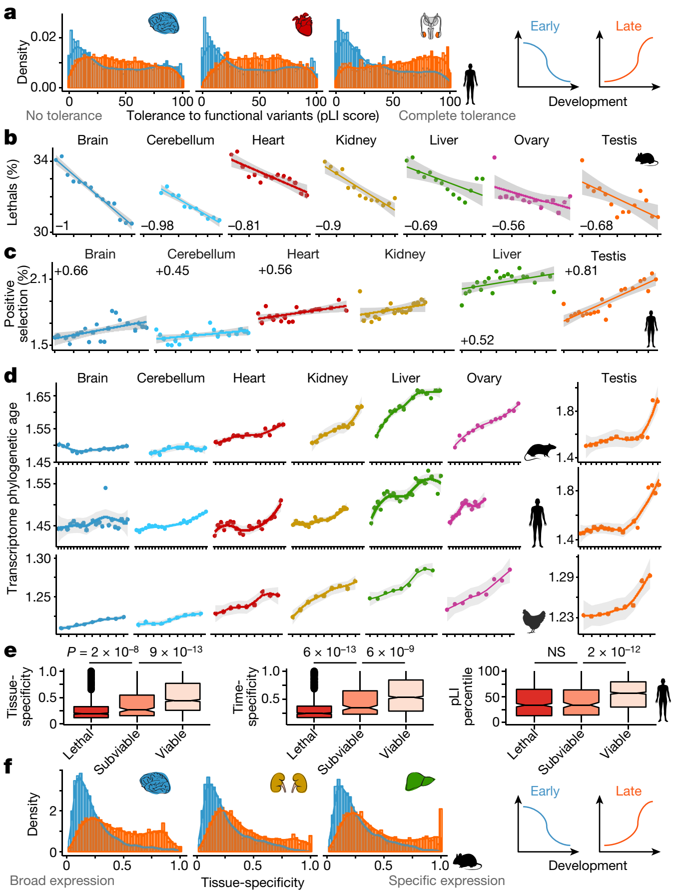


Fig. 3 | Relationships between evolution and development. **a**, Intolerance to functional mutations (probability that a gene is intolerant to a loss-of-function mutation, known as ‘pLI’ score) for human genes whose expression decreases (blue) or increases (orange) during development (4,589/4,478 genes that decrease/increase in brain, 2,673/3,442 in heart, and 2,290/3,794 in testis; all $P < 10^{-10}$, two-sided Wilcoxon rank-sum test). **b**, Percentage of lethal genes expressed at each stage (out of 2,676 knockouts). Weighted average Spearman correlation coefficient is −0.89 ($P = 1 \times 10^{-12}$); all organ-specific Spearman correlations are significant ($P \leq 0.04$). **c**, Percentage of positively selected genes expressed at each stage (out of 13,362 genes tested for positive selection). Ovary excluded owing to lack of postnatal data. Weighted average Spearman correlation coefficient is +0.57 ($P = 5 \times 10^{-11}$); all organ-specific correlations are significant ($P \leq 0.05$). **d**, Phylogenetic age of the transcriptomes of organs throughout development for rat ($n = 18{,}695$ genes), human ($n = 18{,}820$) and chicken ($n = 15{,}155$). Higher values indicate larger contributions of lineage-specific genes (that is, younger transcriptomes). Weighted average Spearman correlation coefficients are +0.87 ($P = 1 \times 10^{-12}$) for rat, +0.77 ($P = 1 \times 10^{-12}$) for human and +0.96 ($P = 1 \times 10^{-12}$) for chicken. All Spearman correlations are significant except for rat brain and cerebellum (ρ: +0.53 to +0.99, P: 0.03–1×10^{-15}). Testis plotted separately because of the young age of sexually mature transcriptomes. **e**, Tissue-specificity, time-specificity (median across organs) and intolerance to functional mutations (pLI) of human orthologues of mouse genes identified as lethal, subviable and viable ($n = 2{,}686$; two-sided Wilcoxon rank-sum test). NS, not significant. Box plots depict the median ± 25th and 75th percentiles, with whiskers at 1.5 times the interquartile range. **f**, Tissue-specificity for mouse genes of which the expression decreases (blue) or increases (orange) during development (3,981/5,164 genes that decrease/increase in brain, 4,631/5,051 in kidney, and 4,270/4,026 in liver; all $P < 10^{-15}$, two-sided Wilcoxon rank-sum test). In **b**–**d**, the *x* axes show samples ordered by stage (Fig. 1a). In **b** and **c**, lines were estimated by linear regression; in **d** by LOESS regression. In **b**–**d**, the 95% confidence interval is shown in grey.

Data Fig. 3b). We recapitulated the Carnegie stage correspondences (rabbit is shifted 1–2 days; Methods; Extended Data Fig. 3a) and found, for example, that gene expression in a newborn opossum is closest to a mouse at embryonic day (E)11.5, matching previous estimates[16].

Organ development includes periods of greater transcriptional change[17]. We identified and characterized these periods across species using our stage correspondences. These periods occur at similar stages across species and are enriched with orthologous genes ($P \leq 0.001$, hypergeometric test; Fig. 2b; Extended Data Figs. 4, 5). In somatic organs, there are two main periods of transcriptional change. The first occurs during embryonic development and is defined by an increase in expression of genes with early organ-specific functions and a decrease in expression of genes involved in cell division and general morphogenesis (Fig. 2b; Extended Data Figs. 4, 5; Supplementary Table 13). The second occurs either postnatally or around birth, depending on the level of maturity of the newborns of the different species. Mammals exhibit great variability in their level of independence at birth, being classified as altricial (born less mature) or precocial (more mature)[6]. These classifications are recapitulated by the developmental transcriptomes, with the altricial newborn opossum at one end of the maturation spectrum and the precocial rhesus macaque at the other (Fig. 2a). This second period of greater transcriptional change is defined by an increase in the expression of genes with late organ-specific functions and, again, by a decrease in expression of genes involved in cell division and morphogenesis (Fig. 2b; Extended Data Figs 4, 5; Supplementary Table 13). Thus, whereas in altricial species this period of intense organ maturation occurs postnatally, in precocial species it overlaps with birth.

Developmental programs can change through shifts in the timing of events—that is, ‘heterochrony’[1]. If the development of an organ were to be shifted in one species, the developmental correspondences for that organ would be different from the global correspondences. Overall, organ-specific correspondences are consistent with the global correspondences, except for early heart development in opossum and early ovary development in human and rabbit (Extended Data Fig. 6; Methods).

However, heterochronies do not have to involve whole organs, they can occur in developmental modules within organs. Indeed, heterochronies occur during the production of gametes[18], a process dependent on meiosis. *Stra8* is the gatekeeper for germ cell entry into meiosis and its role is conserved across vertebrates[3,19]. Therefore, we analysed the expression of *Stra8* and other meiotic genes to identify the start of meiosis in each species, and identified differences in its timing across species (Fig. 2a; Extended Data Fig. 7a–d). During oogenesis, meiotic genes are expressed as early as during embryonic development (mouse), at the boundary between embryonic and fetal development (rat and human), or during late fetal development (rabbit) (Fig. 2a; Extended Data Figs. 6, 7a, b). Although less frequent, we also identified heterochronies in the onset of meiosis during spermatogenesis (Fig. 2a; Extended Data Fig. 7c, d). In spermatogenesis the onset of meiosis marks the beginning of marked changes in cellular composition[20], which sharply change testis transcriptomes (Extended Data Fig. 7e). Starting at birth in rodents and later in rabbit there are also profound changes in ovary transcriptomes (Extended Data Fig. 7e), coinciding with the breakdown of germ cell nests and follicle assembly[21]. Heterochronies are therefore abundant during mammalian gonadal development, representing another mechanism underlying the extreme variability of gonadal morphogenesis[3].

Relationships between evolution and development

After the phylotypic period, the most conserved embryonic stage, morphological differences between species increase as development progresses—von Baer’s divergence[4,5] (Extended Data Fig. 8a). Previous studies assessed the relationship between development and molecular divergence for whole embryos and found that molecular divergence increases as development progresses[22–25]. We recapitulated this observation for individual organs, consistently finding transcriptome correlations between species to decline with developmental time (Extended Data Fig. 8b).

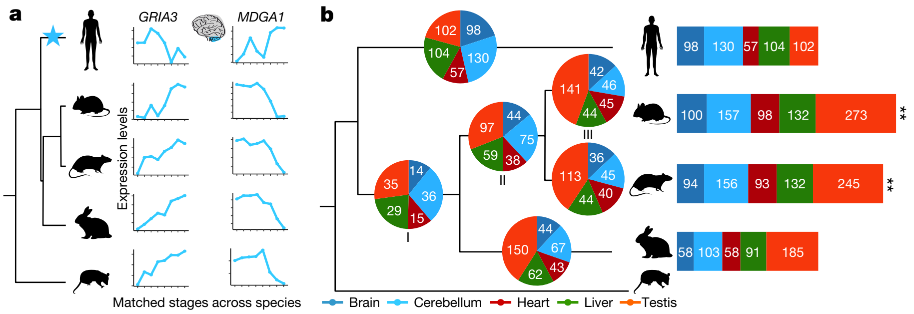


Fig. 4 | Evolution of developmental trajectories. a, Example of two genes that evolved new trajectories in the human cerebellum. *GRIA3* is a glutamate receptor associated with intellectual disability. *MDGA1* encodes a cell surface glycoprotein important for the developing nervous system. **b**, Pie charts depict the number of genes in each organ that evolved new trajectories in each phylogenetic branch (3,980 genes tested in brain, 3,064 in cerebellum, 1,871 in heart, 2,284 in liver and 3,191 in testis). Bar charts depict the total number of trajectory changes in each species. For mouse, that meant adding the changes that occurred at the base of the glires (I), those shared by mouse and rat (II) and those that are mouse specific (III). $^{**}P < 10^{-10}$, binomial test.

Two non-mutually exclusive hypotheses can account for the increasing molecular and morphological divergence during development[26]. First, early development could be under greater functional constraints and be more refractory to change. Second, divergence could be driven by adaptive changes, which are more likely to occur later in development, when the influence of the environment is stronger[26]. To assess potential differences in functional constraints across development, we compared the tolerance to functional mutations of genes used in early versus late development. Across all organs, genes used early in development are less tolerant to loss-of-function mutations ($P < 10^{-10}$, two-sided Wilcoxon rank-sum test; Fig. 3a; Extended Data Fig. 8c). Consistently, using a set of neutrally ascertained mouse knockouts[27], we observed for all organs that the percentage of expressed genes associated with lethality decreases during development (Fig. 3b; Extended Data Fig. 8d). Both observations suggest that early development is subject to stronger functional constraints. Next, we evaluated the relationship between adaptation and development by examining the temporal expression of genes identified as carrying signatures of positive selection[28]. Organs differ in the proportion of positively selected genes: it is highest in liver and testis and lowest in brain tissues (Fig. 3c). However, across all organs, the fraction of expressed positively selected genes increases during development (Fig. 3c). Thus, an increase in positive selection probably also contributes to the progressive molecular and morphological divergence of organs during development.

Organ transcriptomes can also diverge between species owing to species-specific genes[29]. Therefore, we investigated how the contribution of recent gene duplications changes throughout development. For each stage, we calculated an index that combines the phylogenetic age of genes with their expression, in which higher values correspond to younger transcriptomes (Methods). We identified differences between organs similar to those observed for positively selected genes: liver has the youngest developmental transcriptomes, brain tissues the oldest (Fig. 3d). However, across organs, transcriptomes become younger during development, indicating that gene duplications have increasingly more prominent roles (Fig. 3d).

Together, these analyses suggest that the increase in morphological and molecular divergence observed between species during development is driven by a decrease in functional constraints as development advances (Fig. 3a, b), and by a concurrent increase in positive selection (Fig. 3c) and addition of new genes (Fig. 3d).

Pleiotropy and the evolution of development

The breadth of expression across tissues and time points (which we refer to here as pleiotropy) has an influence on gene evolution[30,31]. Therefore, we calculated the tissue- and time-specificity of each gene across development (Extended Data Fig. 9a; Methods; Supplementary Tables 3–9). Time- and tissue-specificity are highly correlated: tissue-specific genes are more likely to be expressed in narrower time windows and vice versa (Pearson correlation coefficients, $r = 0.63$–0.89, $P < 10^{-15}$). Genes also tend to have similar temporal breadths across organs ($r = 0.48$–0.92, $P < 10^{-15}$). As described[32,33], pleiotropy correlates with levels of functional constraint: the more broadly expressed, the more intolerant genes are to functional variation ($r = 0.29$, $P < 10^{-15}$; Extended Data Fig. 9b). Consistently, genes associated with lethality[27] are more broadly expressed than genes associated with subviability, which in turn are more broadly expressed than genes associated with viability (all $P \leq 2 \times 10^{-8}$, two-sided Wilcoxon rank-sum test; Fig. 3e; Extended Data Fig. 9c). This contrasts with measures of tolerance to functional mutations, which distinguish genes associated with lethality or subviablity from viability ($P = 2 \times 10^{-12}$), but do not differentiate between lethality and subviability (Fig. 3e; Extended Data Fig. 9d). Expression pleiotropy is therefore uniquely correlated with phenotypic impact.

Pleiotropy has been proposed as an explanation for the conservation of the phylotypic period[24,34] and is a determinant of the types of mutation that are permissible under selection[30,31]. Therefore, we tested for differences in pleiotropy between genes used early versus late in development and found that genes used earlier have broader spatial and temporal expression than genes employed later (all $P < 10^{-6}$, two-sided Wilcoxon rank-sum test; Fig. 3f; Extended Data Fig. 9e). Because a decrease in pleiotropy can explain both a decrease in functional constraints and an increase in adaptation[30,31], we suggest that it may be a major contributor to the increase in morphological and molecular divergence observed between species during development.

Evolution of developmental trajectories

Differences between species in organ development are often correlated with changes in gene expression. Consequently, we sought to identify genes that evolved new developmental trajectories. Hence, we compared, within a phylogenetic framework, the temporal profiles of orthologous DDGs and identified those with trajectory changes between species (Fig. 4a, b; Supplementary Tables 14–18; Methods).

Brain exhibits the smallest percentage of trajectory changes (11% DDGs), liver and testis the highest (23% and 27%, respectively; Extended Data Fig. 10a). These organ differences are consistent with those observed for positively selected genes and for gene duplications. Thus, across all evaluated mechanisms of evolutionary change, the brain is the slowest evolving organ, whereas liver and testis are the fastest.

In mouse, rat and rabbit the distribution of trajectory changes among organs is similar (Extended Data Fig. 10b). Compared to these species, humans display an excess of trajectory changes in brain (20% changes in human versus 12–13% in others; $P = 1 \times 10^{-5}$, binomial test) and cerebellum (26% in human versus 21–22% in others; $P = 0.02$), and a paucity in testis (21% in human versus 34–37% in others; $P = 1 \times 10^{-10}$) (Extended Data Fig. 10b). Although it is tempting to invoke adaptation, the excess of changes in the human brain tissues could partly stem from differences in sampling (Methods). Overall, rodents evolved a higher number of trajectory changes when compared to human and rabbit ($P < 10^{-10}$; Fig. 4b).

Orthologues tested for trajectory changes are more pleiotropic than the full set of genes in each species, which also includes lineage-specific genes (all $P < 10^{-12}$, two-sided Wilcoxon rank-sum test; Extended Data Fig. 10c). However, among those tested, genes with new trajectories are as pleiotropic as genes with conserved trajectories (Extended Data Fig. 10d). Notably, although genes with trajectory changes are broadly expressed, the changes themselves are organ-specific (Extended Data Fig. 10e). Trajectory changes are restricted to one organ in 93–96% of the cases. This is consistent with the underlying mutations affecting regulatory elements, which control a subset of the total spatiotemporal profile of each gene, and with evolutionary theory, as mutations that affect several organs are less likely to fix in populations[30,31]. However, not all trajectory changes are directly due to regulatory mutations; they can also be caused by changes in cellular composition.

Discussion

We profiled the development of seven major organs, from early organogenesis to adulthood, across multiple mammals, to create an extensive resource (http://evodevoapp.kaessmannlab.org). We used developmental transcriptomes to match stages across species and identified extensive heterochronies during gonadal development. We found the evolution of mammalian organs to be inextricably linked to their development. Organs are most similar between species early in development and then become increasingly more distinct, which is probably explained by changes in pleiotropy during development. Early in development, active genes tend to function in many organs and stages, rendering them more refractory to change. As organs differentiate and mature, active genes have more restricted spatiotemporal profiles, which may reduce functional constraints and facilitate evolutionary change. The increase in species divergence as development progresses has also been described for mammalian limb development[35] and whole embryos[4,5] and we suggest it occurs in developmental systems in which pleiotropy decreases as a function of time.

A next challenge will be to identify the molecular drivers of the new developmental trajectories, which may underlie the evolution of organ phenotypes. It will be important to assess the extent to which these trajectory changes are caused by changes in gene regulation and/or cellular composition. This can be accomplished by complementing the data and results of this study with single-cell transcriptomic and epigenomics datasets and bioinformatic deconvolution procedures. Such endeavours will further advance our understanding of the genetic and developmental foundations of the evolution of mammalian phenotypes.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1338-5.

Received: 6 April 2018; Accepted: 31 May 2019;
Published online 26 June 2019.

1. Pantalacci, S. & Semon, M. Transcriptomics of developing embryos and organs: a raising tool for evo–devo. *J. Exp. Zool. Mol. Dev. Evol.* **324**, 363–371 (2015).
2. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The cellular and molecular landscapes of the developing human central nervous system. *Neuron* **89**, 248–268 (2016).
3. DeFalco, T. & Capel, B. Gonad morphogenesis in vertebrates: divergent means to a convergent end. *Annu. Rev. Cell Dev. Biol.* **25**, 457–482 (2009).
4. Abzhanov, A. von Baer's law for the ages: lost and found principles of developmental evolution. *Trends Genet.* **29**, 712–722 (2013).
5. Kalinka, A. T. & Tomancak, P. The evolution of early animal embryos: conservation or divergence? *Trends Ecol. Evol.* **27**, 385–393 (2012).
6. Ferner, K., Schultz, J. A. & Zeller, U. Comparative anatomy of neonates of the three major mammalian groups (monotremes, marsupials, placentals) and implications for the ancestral mammalian neonate morphotype. *J. Anat.* **231**, 798–822 (2017).
7. Dickinson, M. E. et al. High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508–514 (2016).
8. Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S. & Goldstein, D. B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* **9**, e1003709 (2013).
9. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
10. Cassa, C. A. et al. Estimating the selective effects of heterozygous protein-truncating variants from human exome data. *Nat. Genet.* **49**, 806–810 (2017).
11. Ruderfer, D. M. et al. Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat. Genet.* **48**, 1107–1111 (2016).
12. Hill, M. A. *Embryology Carnegie Stage Comparison* https://jeltsch.org/carnegie_stage_comparison (2017).
13. de Bakker, B. S. et al. An interactive three-dimensional digital atlas and quantitative database of human development. *Science* **354**, aag0053 (2016).
14. Kerwin, J. et al. The HUDSEN Atlas: a three-dimensional (3D) spatial framework for studying gene expression in the developing human brain. *J. Anat.* **217**, 289–299 (2010).
15. Butler, H. & Juurlink, B. H. J. *An Atlas for Staging Mammalian and Chick Embryos* (CRC Press, 1987).
16. Smith, K. K. Early development of the neural plate, neural crest and facial region of marsupials. *J. Anat.* **199**, 121–131 (2001).
17. Dillman, A. A. et al. mRNA expression, splicing and editing in the embryonic and adult mouse cerebral cortex. *Nat. Neurosci.* **16**, 499–506 (2013).
18. Glucksmann, A. Sexual dimorphism in mammals. *Biol. Rev. Camb. Philos. Soc.* **49**, 423–475 (1974).
19. Feng, C. W., Bowles, J. & Koopman, P. Control of mammalian germ cell entry into meiosis. *Mol. Cell. Endocrinol.* **382**, 488–497 (2014).
20. Soumillon, M. et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* **3**, 2179–2190 (2013).
21. Ungewitter, E. K. & Yao, H. H. How to make a gonad: cellular mechanisms governing formation of the testes and ovaries. *Sex Dev.* **7**, 7–20 (2013).
22. Roux, J. & Robinson-Rechavi, M. Developmental constraints on vertebrate genome evolution. *PLoS Genet.* **4**, e1000311 (2008).
23. Kalinka, A. T. et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814 (2010).
24. Hu, H. et al. Constrained vertebrate evolution by pleiotropic genes. *Nat. Ecol. Evol.* **1**, 1722–1730 (2017).
25. Hazkani-Covo, E., Wool, D. & Graur, D. In search of the vertebrate phylotypic stage: a molecular examination of the developmental hourglass model and von Baer's third law. *J. Exp. Zool. B Mol. Dev. Evol.* **304B**, 150–158 (2005).
26. Garfield, D. A. & Wray, G. A. Comparative embryology without a microscope: using genomic approaches to understand the evolution of development. *J. Biol.* **8**, 65 (2009).
27. Koscielny, G. et al. The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Res.* **42**, D802–D809 (2014).
28. Kosiol, C. et al. Patterns of positive selection in six mammalian genomes. *PLoS Genet.* **4**, e1000144 (2008).
29. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **20**, 1313–1326 (2010).
30. Stern, D. L. Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000).
31. Carroll, S. B. Evolution at two levels: on genes and form. *PLoS Biol.* **3**, e245 (2005).
32. Duret, L. & Mouchiroud, D. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol. Biol. Evol.* **17**, 68–74 (2000).
33. Winter, E. E., Goodstadt, L. & Ponting, C. P. Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. *Genome Res.* **14**, 54–61 (2004).
34. Galis, F. & Metz, J. A. Testing the vulnerability of the phylotypic stage: on modularity and evolutionary conservation. *J. Exp. Zool.* **291**, 195–204 (2001).
35. Sears, K., Maier, J. A., Sadier, A., Sorensen, D. & Urban, D. J. Timing the developmental origins of mammalian limb diversity. *Genesis* **56**, e23079 (2018).
36. Plant, T. M., Ramaswamy, S., Simorangkir, D. & Marshall, G. R. Postnatal and pubertal development of the rhesus monkey (*Macaca mulatta*) testis. *Ann. NY Acad. Sci.* **1061**, 149–162 (2005).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Ethics statement. The human prenatal samples were provided by the MRC-Wellcome Trust Human Developmental Biology Resource (HDBR) and are from elective abortions with normal karyotypes. The tissue donations were made entirely voluntarily by women undergoing termination of pregnancy. Donors were asked to give explicit written consent for the fetal material to be collected, and only after they had been counselled about the termination of their pregnancy. The human postnatal samples were provided by the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland (USA) and by the Chinese Brain Bank Center (CBBC) in Wuhan (China). They originated from individuals with diverse causes of death that, given the information available, were not associated with the organ sampled. Written consent for the use of human tissues for research was obtained from all donors or their next of kin by the respective tissue banks. The rhesus macaque samples were provided by the Suzhou Experimental Animal Center (China). All rhesus macaques used in this study suffered sudden deaths for reasons other than their participation in this study and without any relation to the organ sampled. The use of all samples for the work described in this study was approved by an ERC Ethics Screening panel (associated with H.K.'s ERC Consolidator Grant 615253, OntoTransEvol) and local ethics committees in Lausanne (authorization 504/12) and Heidelberg (authorization S-220/2017).

Human developmental samples. We started sampling human prenatal development at 4 weeks post-conception (WPC) and then sampled the developing organs each week until 20 WPC (except for 14, 15 and 17 WPC). There are no samples available between 20 and 38 WPC. Postnatally we sampled neonates, 'infants' (6–9 months), 'toddlers' (2–4 years), 'school' (7–9 years), 'teenagers' (13–19 years), and then adults from each decade until 65 years of age (Supplementary Tables 1, 2). Human ovary development was only sampled prenatally (until 18 WPC) and kidney development was sampled up until (and including) 8 years of age ('school'). Prenatally, we considered as biological replicates individuals from the same developmental week. Hence, for example, individuals from Carnegie stages 13 and 14 were considered to be replicates (that is, 4 WPC) even though they were not at the same developmental stage according to phenotypic milestones. Supplementary Table 2 provides the precise age of the donors. The number of biological replicates ranges from 1 to 4 (median of 2), for a total of 297 RNA-sequencing (RNA-seq) libraries.

Other species developmental samples. In mouse (*Mus musculus*, outbred strain CD-1, RjOrl:SWISS), we started sampling prenatal development at E10.5 and then collected samples daily until birth (that is, until E18.5). Postnatally we sampled individuals at 5 stages: postnatal day (P)0, P3, P14, P28 and P63. There are four replicates (two males and two females) per stage, except for ovary and testis, in which we aimed for two replicates, for a total of 316 RNA-seq libraries.

In rat (*Rattus norvegicus*, outbred strain Holtzman SD), the time series started at E11 and covered prenatal development daily until birth (that is, until E20). Postnatally we sampled individuals at 6 stages: P0, P3, P7, P14, P42 and P112. We generated replicates as described for mouse, for a total of 350 RNA-seq libraries.

In rabbit (*Oryctolagus cuniculus*, outbred New Zealand breed), we started sampling prenatal development at E12 and then sampled 11 time points up until (and including) E27 (gestation length is approximately 29–32 days). Postnatally we sampled individuals at 4 stages: P0, P14, P84 and between P186 and P548. The number of replicates is the same as described for the rodents, for a total of 315 RNA-seq libraries.

For rhesus macaque (*Macaca mulatta)*, the sample collection started at a fetal stage (E93) and we sampled five stages before birth (until E130; gestation lasts around 167 days). Postnatally we sampled eight stages selected to match the human time series (Supplementary Tables 1, 2). The number of replicates ranges from 1 to 4 (median of 2), for a total of 168 RNA-seq libraries. Ovary development was not sampled.

Because opossums (*Monodelphis domestica*) are born in a very immature state[6], the stages of organ development sampled in the other species during prenatal development occur postnatally in the marsupial. Accordingly, we sampled one prenatal stage (E13.5; gestation length is ~14–15 days) and then sampled 14 postnatal stages, more densely right after birth and then at increasingly longer intervals (Supplementary Table 1). There is a median of three biological replicates per stage for somatic organs and two for the gonads, for a total of 232 RNA-seq libraries.

In chicken (*Gallus gallus*, the red junglefowl, the progenitor of domestic chicken), we started sampling organ development at a fetal stage (E10) and sampled three additional stages until E17 (egg incubation lasts approximately 21 days). We then sampled postnatal development at five stages: P0, P7, P35, P70 and P155. There is a median of four biological replicates per stage for somatic organs (two males and two females) and two for the gonads, for a total of 215 RNA-seq libraries.

This resource consists of 1,893 libraries, covering the development of 7 organs, 9–23 developmental stages (depending on the species) and a median of 2–4 replicates per stage (full details in Supplementary Tables 1, 2).

Organ dissections. Mammalian embryos are morphologically similar[4], and this similarity extends to the internal organs. Early in development, the only clear morphological difference between the organs of the different species, when present, is size. We started collecting samples when the organs could be dissected and isolated from nearby tissues. For the brain, this was possible across the entire time series. Human and rhesus macaque prenatal brain was divided into two regions: forebrain together with the midbrain (referred to as the 'brain') and hindbrain (referred to as the 'cerebellum'). Human and rhesus macaque postnatal brain and cerebellum samples comprise the dorsolateral prefrontal region of the cerebral cortex and lateral cerebellar cortex, respectively. For the other species, the dissected brain samples correspond to the cerebral hemispheres (without the olfactory bulbs). The early cerebellum samples correspond to the prepontine hindbrain-enriched brain region (until the period matching a mouse E15.5) and from E16.5 onwards only to the cerebellum. For mouse, rat and rabbit, the earliest developmental samples consist of whole brains, which were analysed as part of the brain time series (Supplementary Table 1). We dissected heart samples across the entire time series. At the earliest stage sampled, the heart is beating and the four chambers are already present[37]. For most species, the liver could also be individually dissected at the start of the time series. The developing gonads are visible as a paired structure on the ventromedial surface of the mesonephros before the start of the time series. However, depending on the species, we were only able to completely isolate the developing gonads at later stages. The same was true for the developing kidneys. In chicken only the left ovary develops and this was the one dissected. The dissections include the main organ structures/cell types in all species but, with the exception of the early samples, they did not include the whole organ.

RNA extraction and sequencing. RNA was extracted using the RNeasy protocol from QIAGEN. RNeasy Micro columns were used to extract RNA from small ($<$5 mg) or fibrous samples and RNeasy Mini columns were used to extract RNA from larger samples. The tissues were homogenized in RLT buffer supplemented with 40 mM dithiothreitol (DTT) or QIAzol. To ensure that we were not introducing technical biases by using two different homogenization procedures, we generated libraries for four samples (two adult rat brains and two adult rat testis) using RLT buffer with DTT or QIAzol (with two technical replicates). All libraries from the same organ showed a Pearson correlation coefficient value of $\geq$0.99 irrespective of the homogenization procedure (the median correlation between replicates in the rat dataset was 0.99). RNA quality was assessed using the Fragment Analyzer (Advanced Analytical). The RNA-seq libraries were created using the TruSeq Stranded mRNA LT Sample Prep Kit (Illumina) and sequenced on the HiSeq 2500 platform (multiplexed in sets of 6 or 8). All libraries are strand-specific, 100 nucleotides single-end, and were sequenced to a median depth of 33 million reads at the Lausanne Genomic Technologies Facility (Supplementary Table 2). The sequencing depth was uniform across the libraries (5% and 95% quantiles of 20 and 54 million reads, respectively). A subset of the adult libraries was used in previous publications[38,39].

Quality control of the libraries and estimation of expression levels. We mapped the reads from each library against the species reference genome (Supplementary Table 19) using GSNAP (22-10-2014)[40]. The alignments were guided by the known gene annotations and the discovery of novel splice sites was enabled (Supplementary Tables 19, 20). We used HTSeq (v.0.6.1)[41] to generate read counts for the set of protein-coding genes (Supplementary Tables 19, 20). Only uniquely mapping reads were allowed. We normalized the count data using the method TMM as implemented in the package EdgeR (v.3.14.0)[42]. EdgeR was also used to generate the expression tables used in the study. Expression levels were calculated as counts per million (CPM) or in reads per kilobase of exon model per million mapped reads (RPKM). The alignment files were manipulated using samtools (v.0.1.18)[43] and general alignment statistics created using Picard (1.86)[44] (Supplementary Table 20).

We used the RNA quality number (RQN) of Fragment Analyzer to evaluate the quality of the samples and generally generated sequencing data for those with high values ($\geq$7). However, because we also sequenced libraries with lower RQN values, we performed an additional check on RNA integrity after sequencing. We used the 'CollectRnaSeqMetrics' Picard tool to calculate the distribution of read coverage along transcripts. RNA degradation leads to a bias in read coverage by favouring the $3'$ end of genes and this can be identified by calculating the median $3'$ bias of transcript coverage. We excluded from our dataset all libraries that showed a significant $3'$ bias in read coverage.

We evaluated the quality of the sequenced libraries using unsupervised hierarchical clustering (hclust) and PCA (FactoMineR 1.34[45]) as implemented in R[46]. In a PCA, the developmental samples from an organ of a given species should be ordered by developmental time in a characteristic U or V shape[47]. Samples with low RNA quality, insufficient sequencing depth, or showing potential contamination with other tissues appeared as outliers in the organ PCAs and were excluded (the outlier status was confirmed using hierarchical clustering). The global and organ-specific PCAs used as input the read counts after applying the variance stabilizing transformation (VST) implemented in DESeq2 (v.1.12.4)[48]. The sex of the samples was confirmed using the female-specific genes *Xist* (eutherians), *Rsx* (opossum) and *CDC34* (chicken) (and for eutherians with available Y chromosomes also with the Y-linked gene *Ddx3y*) using Bedtools (v.2.18)[49]. Finally, we removed from the dataset libraries in which the correlation among replicates (Spearman's ρ) was lower than 0.90. We are making available the libraries that passed the general quality control but had correlations with their replicates <0.90, but they were not used in this study and are marked as such in Supplementary Table 2.

DDGs. In each organ, we identified the genes with dynamic temporal profiles (DDGs) using maSigPro, an R package designed for transcriptomic time courses[50,51]. We used as input the count tables from EdgeR (in CPM), and only excluded genes that did not reach a minimum of 10 reads in at least 3 libraries. We ran maSigPro on the log-transformed time (measured in days post-conception) with a degree $= 3$ (polynomial). We considered genes as DDGs in an organ when the goodness-of-fit (R^2) was at least 0.3 and the maximum RPKM in that organ was at least 1. The lists of DDGs in each organ and species are provided in Supplementary Tables 3–9.

We identified differences between species and organs in the number of DDGs (Extended Data Fig. 2a). However, technical aspects of the datasets can explain these differences, particularly those between species. First, owing to the nature of the statistical test used, the power to call differential temporal expression depended on the magnitude of the expression change and on the agreement between the biological replicates. Smaller expression changes could only be detected if there was strong agreement between the biological replicates. There are differences between species in the median correlation across replicates (Spearman's ρ: 0.94–0.99) and these are strongly correlated with the number of DDGs detected ($\rho = 0.66$, $P < 10^{-6}$). Two factors contribute to the species differences in the correlations among replicates. One is the amount of genetic diversity (for example, lower in mouse than human); the other is how close biological replicates are in terms of development. In rodents the biological replicates are from identical developmental stages (sometimes even the same litter) but in primates the biological replicates span developmental periods. Second, there are differences between species and organs in the length of the time series (Supplementary Table 1). Notably, in chicken and rhesus macaque, we are missing the earliest developmental stages, when key developmental processes occur. Finally, some differences could also derive from differences in genome annotation.

We characterized DDGs using three different metrics of functional constraint: (1) the residual variation intolerance score (RVIS); (2) the probability of being intolerant to loss-of-function mutations (pLI score); and (3) the selection against heterozygous loss of gene function (s_{het}). All metrics were applied to data from the Exome Aggregation Consortium (ExAC)[9]. We obtained the pLI and RVIS scores from ref. [7] and s_{het} values from ref. [10]. We also used the copy number variation (CNV) intolerance score as applied to the ExAC data from ref. [11]. The lists of transcription factors were from the animalTFDB (v.2.0)[52].

Stage correspondences across species. We identified stage correspondences across species using the set of 1:1 DDGs in all species. Because of the shorter time series, we did not require genes to be DDGs in rhesus macaque. We used the combined information from the somatic organ DDGs to calculate the Spearman correlations between all stages in mouse and all stages in each of the other species (using for each stage the median across replicates). We then ran the dynamic time warping algorithm implemented in the R package 'dtw' (v.1.18-1)[53] to identify the optimal alignment between each of the two time series. We ran dtw using as step pattern 'symetricP05' (except for rhesus macaque and chicken where the late fetal start required us to use 'asymmetric' with 'begin.open = T'). When a stage in a given species matched two or more stages in mouse, we kept the one with the highest correlation (Extended Data Fig. 3a, b). Our cross-species stage correspondences recapitulated the stage correspondences based on the Carnegie staging for all species except rabbit (shifted 1–2 days; Extended Data Fig. 3a). An independent, neural development-based stage assignment across mammals[54], suggested an even more pronounced shift (3–4 days) forward for rabbit.

We then evaluated whether the stage correspondences based on the combined information from the somatic organs were consistent with the information available for each individual organ. For each organ and stage in mouse, we selected in the other species the stage with the maximum correlation plus all stages within 1% of the maximum correlation. We then fitted a local polynomial regression (LOESS) to identify the organ-specific correspondences (Extended Data Fig. 6). Overall, the global stage correspondences are within the 98% confidence interval of organ-specific correspondences, suggesting that a single stage correspondence can be used for all organs. But there are exceptions. The heart-specific correspondence between mouse and opossum differs from the global correspondence early in development, suggesting that in relation to the other organs, heart development in opossum could be shifted forward, that is, be in a more advanced developmental stage. Early opossum development is characterized by heterochronies in the craniofacial, axial and limb skeleton that allow the neonates to crawl without their mother's help to the teat immediately after birth[16,55]. It is possible that heart development is also shifted forward to accommodate the greater demands of what is postnatal life in opossum, and still prenatal life in the other species. The other potential exception applies to early ovary development in human and rabbit, where we observe development to be shifted forward in the two species. Using the ovary-specific correspondences, the heterochronies associated with the onset of meiosis during oogenesis in these species are even more pronounced than when using the global stage correspondences (Extended Data Fig. 6).

We were underpowered to detect shifts in individual organs that encompass a small number of adjacent time points. Throughout organ development, the correlation between adjacent stages is, as expected, high, and we would only be able to detect small shifts if they led to a high discordance between species (that is, significantly lower correlations for a short interval when compared to the rest of the time series). The only instance of this in our dataset was during testis development, in association with the onset of meiosis (inset in Extended Data Fig. 6). The onset of meiosis marks the beginning of notable changes in cell composition in the testis[20], which make the transcriptomes that flank this event distinct from each other (Extended Data Fig. 7e), thereby allowing the detection of significant differences between species between adjacent stages.

Periods of greater transcriptional change. For each species, we identified the genes that are differentially expressed between adjacent time points (based on the cross-species stage correspondences) using DESeq2 (v.1.12.4)[48]. We required the adjusted *P* value to be ≤ 0.05 and the $\log_2$ fold change to be ≥ 0.5. Differences between species in the number of replicates and in the correlation among the replicates (see 'DDGs' section) affected our power to call differential expression. Both factors led to lower power to detect differential expression in primates than in mouse, rat and rabbit. Therefore, we are probably underestimating the amount of transcriptional change in humans.

Relationships between evolution and development. In Fig. 3a, f (and in Extended Data Figs. 8c and 9e), we compare the tolerance to functional mutations and the time- and tissue-specificity of genes used early versus late in development in human and mouse. For each species, we identified these genes in the following way. First, we identified the most common profiles in each organ using the soft-clustering approach (c-means) implemented in the R package mFuzz (v.2.32.0)[56,57]. The clustering was restricted to DDGs and we used as input the read counts after applying the VST to the raw counts implemented in DESeq2 (v.1.12.4)[48]. The number of clusters was set to 6–8 depending on the organ. For each organ, we settled on a cluster number when increasing it would not add a new cluster but instead split a previous cluster in two. We considered that a cluster was split into two when the median profile of the genes in the two new clusters was similar and when functional enrichment analyses were also similar between the clusters. Between 86% and 92% of genes in mouse and 89% and 93% of genes in human were clearly assigned to one of the clusters (cluster membership ≥ 0.7). Among these genes, those assigned to clusters characterized by a decrease in expression during development were classified as genes used early in development and those assigned to clusters with the opposite profile were classified as genes used late in development. Genes assigned to clusters with other profiles were classified as other. The classification of each gene in each organ as 'early', 'late', 'other' or 'NA' (if a gene is not DDG in the organ or if it has a membership <0.7) is provided in Supplementary Tables 3 (mouse) and 6 (human).

In Fig. 3b (and Extended Data Fig. 8d), we used a set of neutrally ascertained mouse knockouts that consists of 2,676 protein-coding genes: 646 are classified as lethal, 257 as subviable (less than 12.5% of expected pups) and 1,773 as viable. These were the data on viability available for download on 7 June 2017 from the International Mouse Phenotyping Consortium (IMPC)[27]. For each developmental stage, the denominator is the number of genes expressed that were tested for lethality and the numerator the genes among those that resulted in a lethal phenotype. In Extended Data Fig. 8d, we also include in the numerator the genes that resulted in a subviable phenotype (top) and exclude from the analysis a set of housekeeping genes previously identified[58] (bottom). We excluded housekeeping genes because they are typically most highly expressed early in development and are enriched among lethals[22].

In Fig. 3c, we used a set of genes previously identified[28] as carrying evidence for coding-sequence adaptation in mammals. For each developmental stage, the denominator is the number of expressed genes that were tested for signatures of positive selection and the numerator is the number of genes among those with evidence for positive selection.

In Fig. 3d, we plotted the age of the transcriptome for each developmental stage. The 'age of the transcriptome' was inspired by the transcriptome age index (TAI) that was previously developed[59] but differs fundamentally from it in that we are dating the emergence of individual genes and not of gene families (that is, the emergence of the founder member of a gene family). The TAI measure is a weighing procedure (weighted arithmetic mean) that gives greater weight to young duplicates. The age of the duplicates was determined based on syntenic alignments across vertebrates and parsimony as previously described[60]. The pipeline was run for human, mouse, rat and chicken (based on Ensembl 69 annotations). Most new genes emerged via small-scale duplications in mammals[60]. Genes predating the vertebrate split were given a score of 1, genes shared by amniotes were given a score of 2 and so on, until genes that are species-specific were given the maximum score. The range of the score differed between species depending on the number of outgroup lineages available (more lineages allowed for more details in the phylogeny) and therefore this index cannot be compared across species, only within species (that is, across organs). The score assigned to each gene was multiplied by the expression of the gene (but only if RPKM > 1). The results reported used the $\log_2$-transformed RPKM values but similar trends were obtained using the raw RPKM values. Higher values indicate larger contributions of recently duplicated genes (that is, younger transcriptomes).

Pleiotropy indexes. The time- and tissue-specificity indexes are based on the Tau metric of tissue specificity[61]. To calculate our tissue-specificity index, we applied the Tau formulation to the maximum expression observed during development in each organ. The time-specificity index uses the Tau formulation for time points instead of organs. Both indexes range from 0 (broad expression) to 1 (restricted expression). These indexes are provided in Supplementary Tables 3–9.

Comparing developmental trajectories. We compared developmental trajectories between human, mouse, rat, rabbit and opossum. Rhesus macaque and chicken were not included because their time series start at a late fetal stage. We used GPClust, a method to cluster time series using Gaussian processes[62–64], to identify the most common developmental trajectories in each organ. We used the expression (vst-counts) of all available orthologous DDGs as the input (median across replicates for matching stages only). We set the noise variance (k2.variance.fix) to 0.7. GPClust assigned each gene the probability of belonging to each of the trajectories (clusters). We then inferred within a phylogenetic framework the probability that there were changes in developmental trajectories, that is, that genes changed their cluster assignment in specific branches. We did this in a two-step approach. First, we inferred ancestral cluster probabilities along the tree by calculating the weighted averages from the child-nodes. The weights are given by the inverse branch lengths, which were retrieved from TimeTree[65], so that closer child-nodes have more weight. To detect changes in the overall pattern at each branching in the tree we calculated the probability that its two nodes are in the same cluster. If the probability was below 1%, we called the node as having changed. Second, after identifying all such nodes, we mapped the change to one of the two branches by comparing the two children of the node with the outgroup node. The results are provided in Supplementary Tables 14–18.

It was not always possible to identify the specific branch where changes occurred. This was either because changes were also detected at neighbouring nodes (making unclear where the change occurred) or because two nodes differed at the threshold used (1%) but they were both not different from their joint closest relative (for example, when a call was made for mouse versus rat but neither for mouse versus rabbit nor rat versus rabbit). These calls are classified as 'NA' in Supplementary Tables 14–18. Finally, changes between opossum and the eutherian species could not be polarized because of the lack of an outgroup (classified as eutherian/opo in Supplementary Tables 14–18). These changes were included in Extended Data Fig. 10. The results for genes that have one trajectory change across the phylogeny are summarized in Fig. 4 and Extended Data Fig. 10.

Differences in developmental trajectories between species can be created by changes in the expression levels of genes in homologous cell populations, by expansions/contractions of homologous cell populations, or by differences in the cell populations that express a given gene (all non-mutually exclusive possibilities). We chose a conservative cut-off (1%) to identify trajectory changes because our aim was to identify those with the largest biological effects. As a consequence, we are probably enriching for differences between species created by abrupt changes in the size of homologous cell populations, differences in the cell populations that express a given gene, and/or by differences in expression levels of genes in homologous cell populations that are time-specific (as opposed to being progressive during development).

The effect of organ complexity on estimates of species divergence. Organ complexity can affect estimates of gene expression. Expression changes in low abundant cell types that can be detected in simpler organs can potentially go undetected in more complex organs[1]. Because the brain has a higher cellular complexity than the other organs studied[66], it may appear to be more conserved between species than it really is. Indeed, we found brain tissues to be consistently the slowest evolving, irrespective of the variable being measured. Developmental datasets can help to address the problem of comparing organs with different levels of complexity. Organs are more homogeneous early in development and then progressively increase in complexity (for example, the number of distinct cell types increases during development)[1]. This means that when we analyse entire time series, we are comparing organs at different levels of complexity, including early in development when organ complexity is lowest. Throughout the entire times series, we consistently observed more similarities between the transcriptomes of species for the brain than for the other organs (Extended Data Fig. 8b), including at the earliest stages. We also observed that overall organs are most similar across species early in development (when the power to identify differences would be greatest), and then progressively diverge through time (Extended Data Fig. 8b). Finally, the differences between organs were also consistent throughout the entire development when evaluating the percentage of expressed positively selected genes (Fig. 3c) and the contribution of recent gene duplications (Fig. 3d). Together, these observations suggest that the observed differences between organs in their evolutionary rates are independent of organ complexity. We could, however, be underestimating the total divergence of organs, particularly in adults.

General statistics and plots. Unless otherwise stated, all statistical analyses and plots were done in R[46]. Plots were created using the R packages ggplot2 (v.2.2.1)[67], gridExtra (v.2.2.1)[68], reshape2 (v.1.4.2)[69], plyr (v.1.8.4)[70], and factoextra (v.1.0.4)[71]. All functional enrichment analyses were done using the R implementation of WebGestalt (v.0.0.5)[72]. All packages and versions used are described in Supplementary Table 20.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw and processed RNA-seq data have been deposited in ArrayExpress with the accession codes: E-MTAB-6769 (chicken), E-MTAB-6782 (rabbit), E-MTAB-6798 (mouse), E-MTAB-6811 (rat), E-MTAB-6813 (rhesus macaque), E-MTAB-6814 (human) and E-MTAB-6833 (opossum) (https://www.ebi.ac.uk/arrayexpress/). The temporal profiles of individual genes across organs and species can be visualized and downloaded using the web-based application: http://evodevoapp.kaessmannlab.org.

37. Bruneau, B. G. Signaling and transcriptional networks in heart development and regeneration. *Cold Spring Harb. Perspect. Biol.* **5**, a008292 (2013).
38. Carelli, F. N., Liechti, A., Halbert, J., Warnefors, M. & Kaessmann, H. Repurposing of promoters and enhancers during mammalian evolution. *Nat. Commun.* **9**, 4066 (2018).
39. Marin, R. et al. Convergent origination of a *Drosophila*-like dosage compensation mechanism in a reptile lineage. *Genome Res.* **27**, 1974–1987 (2017).
40. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
41. Anders, S., Pyl, P. T. & Huber, W. HTSeqa Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
42. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
43. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
44. Picard. http://broadinstitute.github.io/picard (2015).
45. Lê, S., Josse, J. & Husson, F. FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
46. R Core Team. *R: A Language and Environment for Statistical Computing* (2014).
47. Anavy, L. et al. BLIND ordering of large-scale transcriptomic developmental timecourses. *Development* **141**, 1161–1166 (2014).
48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
49. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
50. Nueda, M. J., Tarazona, S. & Conesa, A. Next maSigPro: updating maSigPro Bioconductor package for RNA-seq time series. *Bioinformatics* **30**, 2598–2602 (2014).
51. Conesa, A., Nueda, M. J., Ferrer, A. & Talón, M. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* **22**, 1096–1102 (2006).
52. Zhang, H. M. et al. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res.* **43**, D76–D81 (2015).
53. Giorgino, T. Computing and visualizing dynamic time warping alignments in R: the dtw package. *J. Stat. Softw.* **31**, 1–24 (2009).
54. Clancy, B., Darlington, R. B. & Finlay, B. L. Translating developmental time across mammalian species. *Neuroscience* **105**, 7–17 (2001).
55. Smith, K. K. Craniofacial development in marsupial mammals: developmental origins of evolutionary change. *Dev. Dyn.* **235**, 1181–1193 (2006).
56. Futschik, M. E. & Carlisle, B. Noise-robust soft clustering of gene expression time-course data. *J. Bioinform. Comput. Biol.* **3**, 965–988 (2005).
57. Kumar, L. & E Futschik, M. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2**, 5–7 (2007).
58. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet.* **29**, 569–574 (2013).

59. Domazet-Lošo, T. & Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815–818 (2010).
60. Zhang, Y. E., Vibranovski, M. D., Landback, P., Marais, G. A. & Long, M. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol*. **8**, e1000494 (2010).
61. Yanai, I. et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
62. Hensman, J., Lawrence, N. D. & Rattray, M. Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters. *BMC Bioinformatics* **14**, 252 (2013).
63. Hensman, J., Rattray, M. & Lawrence, N. D. Fast nonparametric clustering of structured time-series. *IEEE Trans. Pattern Anal. Mach. Intell*. **37**, 383–393 (2015).
64. Hensman, J., Rattray, M. & Lawrence, N. D. Fast variational inference in the conjugate exponential family. In *Proc. NIPS'12 Proceedings 25th International Conference on Neural Information Processing Systems* Vol. 2, 2888–2896 (2014).
65. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol*. **34**, 1812–1819 (2017).
66. Vickaryous, M. K. & Hall, B. K. Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol. Rev. Camb. Philos. Soc*. **81**, 425–455 (2006).
67. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2009).
68. Auguie, B. gridExtra: Miscellaneous Functions for "Grid" Graphics. v.2.2.1 (2015).
69. Wickham, H. Reshaping data with the reshape package. *J. Stat. Softw*. **21**, 1–20 (2007).
70. Wickham, H. The split-apply-combine strategy for data analysis. *J. Stat. Softw*. **40**, 1–29 (2011).
71. Kassambara, A. & Mundt, F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. v.1.0.4 (2017).
72. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res*. **45**, W130–W137 (2017).

Acknowledgements We thank R. Arguello, F. Lamanna, I. Sarropoulos, M. Sepp and members of the Kaessmann group for discussion; I. Moreira for developing the Shiny application; M. Sanchez-Delgado and N. Trost for figure assistance; P. Dhakal for assistance collecting rat samples; S. Nef for assistance collecting gonadal samples; the Lausanne Genomics Technology Facility for high-throughput sequencing; the Joint MRC/Wellcome (MR/R006237/1) Human Developmental Biology Resource (HDBR), the Maryland Brain Collection Center and the Chinese Brain Bank Center for providing human samples; and the Suzhou Experimental Animal Center for providing rhesus macaque samples. Computations were performed at the Vital-IT Center of the SIB Swiss Institute of Bioinformatics and at the bwForCluster from the Heidelberg University Computational Center (supported by the state of Baden-Württemberg through bwHPC and the German Research Foundation - INST 35/1134-1 FUGG). M.J.S. was supported by the National Institutes of Health (HD020676); P.J. by the European Research Council (322206, Genewell); Y.E.Z. by the National Natural Science Foundation of China (31771410, 91731302); M.C. by Fundação para a Ciência e Tecnologia through POPH-QREN funds from the European Social Fund and Portuguese MCTES (IF/00283/2014/CP1256/CT0012); S. Anders by the German Research Foundation (SFB 1036). This research was supported by grants from the European Research Council (615253, OntoTransEvol) and Swiss National Science Foundation (146474) to H.K. and by the Marie Curie FP7-PEOPLE-2012-IIF to M.C.-M. (329902).

Author contributions M.C.-M. and H.K. conceived and organized the study based on H.K.'s original design. M.C.-M. performed all analyses. M.C.-M. and H.K. wrote the manuscript, with input from all authors. J.H., D.V., C.S., M.J.S., P.G.F., S. Afonso, M.C., J.M.A.T., J.L.V., A.F., P.J., R.B., S. Lisgo, S. Lindsay, P.K. and J.B. collected samples. J.H., D.V., A.L., K.A. and C.R. performed RNA-seq experiments. B.V. and W.H. contributed to analyses on trajectory changes. S.O., S. Anders and P.V.M. contributed to analyses on developmental correspondences. C.C., Y.S. and Y.E.Z. contributed to analyses on transcriptome age. M.M. and D.N.C. contributed to data interpretation. I.X. organized high-performance computational resources. K.H. organized high-throughput sequencing.

Competing interests The authors declare no competing interests.

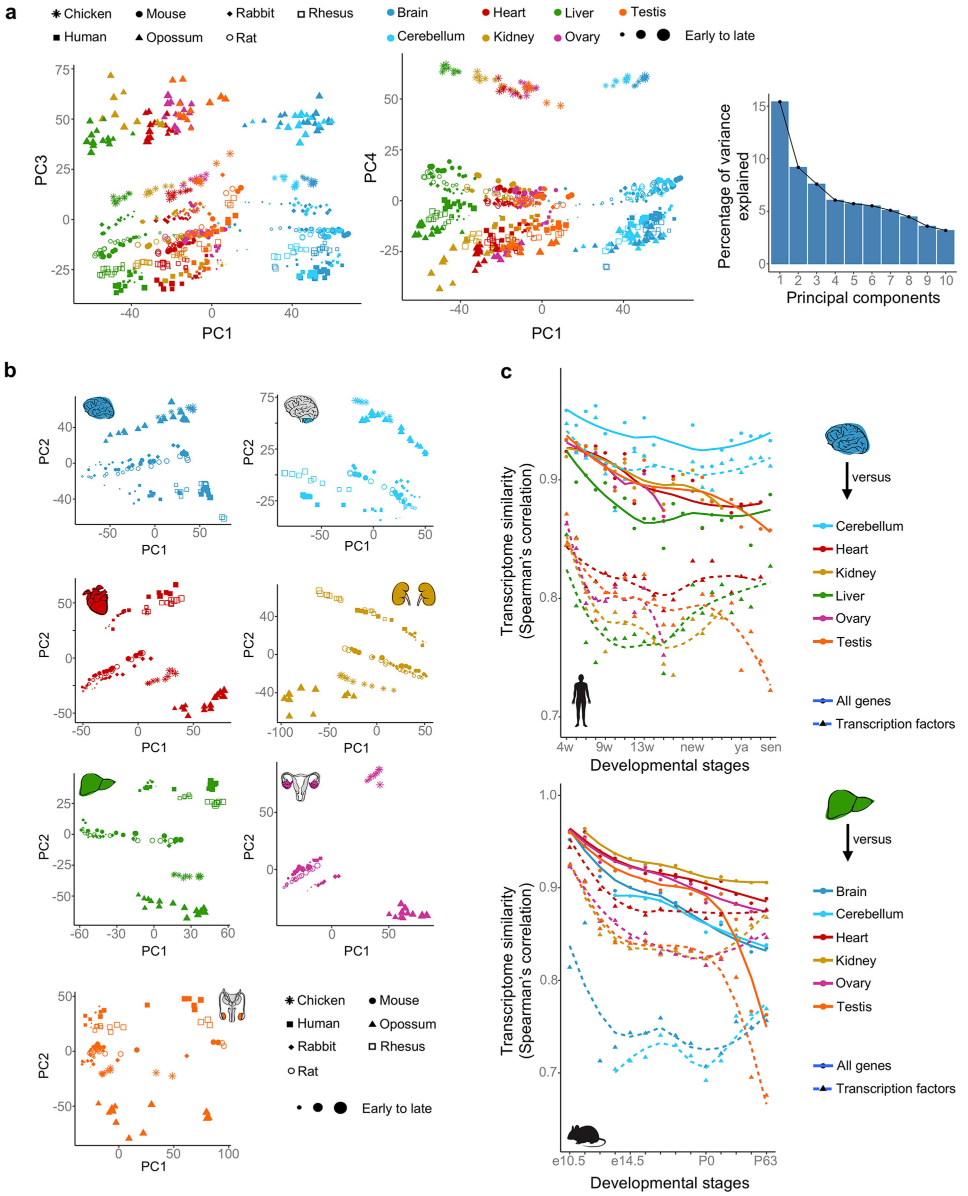
Additional information

Supplementary information. is available for this paper at https://doi.org/10.1038/s41586-019-1338-5.

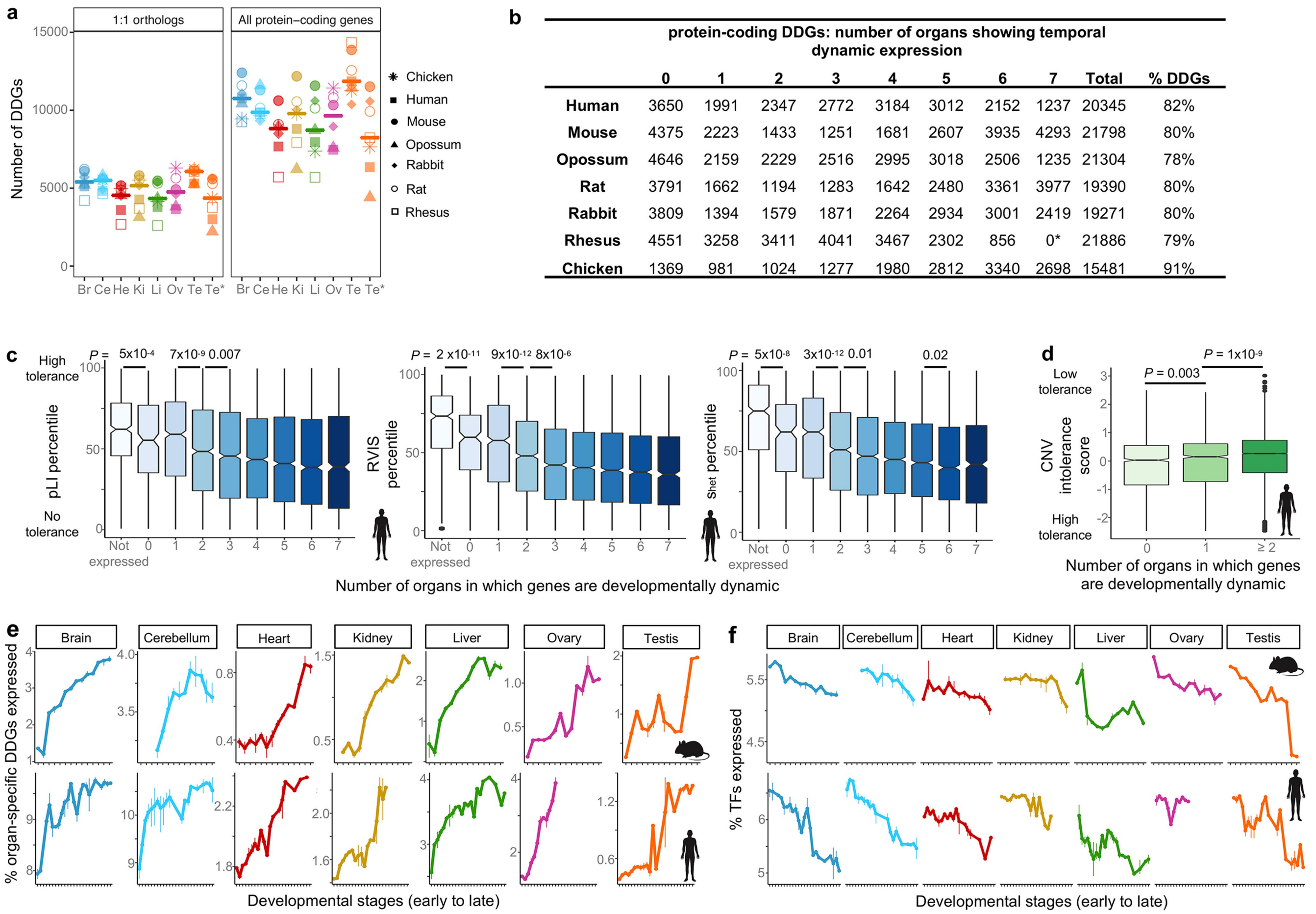
Correspondence and requests for materials should be addressed to M.C. or H.K.

Peer review information *Nature* thanks Pavel Tomancak and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | Organ developmental transcriptomes. a, PC3 and PC4 of the PCA based on 7,696 1:1 orthologues depicted in Fig. 1b (each dot represents the median across replicates), and scree plot describing the amount of variance explained by the first 10 principal components. **b**, PCAs of individual organs ($n = 7,696$ 1:1 orthologues). **c**, Correlation of expression levels throughout development between human brain and the other organs (20,345 genes) (top), and between mouse liver and the other organs (21,798 genes) (bottom). Similar patterns were observed using other organs as the focal organ, and species. For human, the prenatal data are in weeks (w) postnatally; new, newborn; sch, school age (7–9 years); ya, young adult (25–32 years); sen, senior (58–65 years).



protein-coding DDGs: number of organs showing temporal dynamic expression

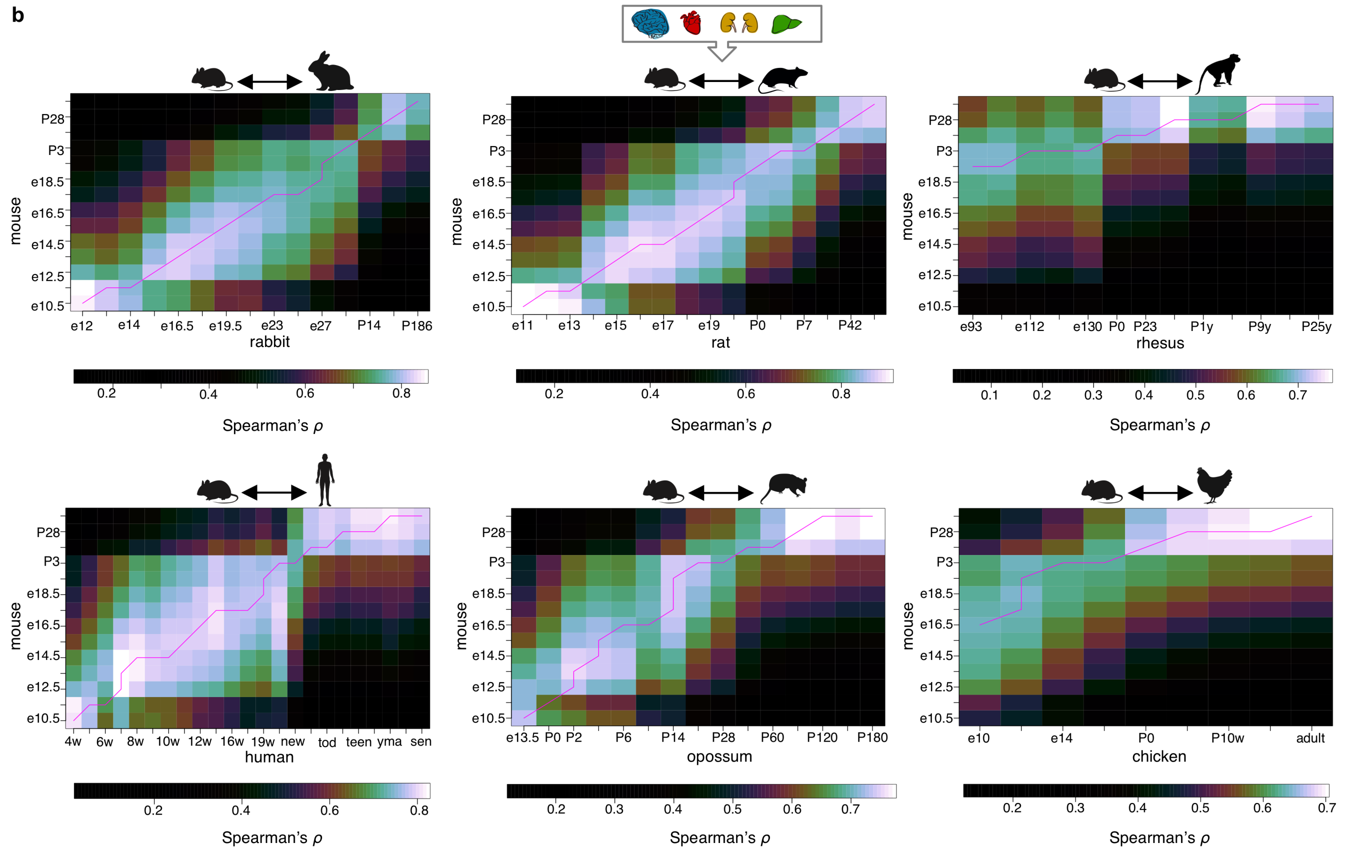
	0	1	2	3	4	5	6	7	Total	% DDGs
Human	3650	1991	2347	2772	3184	3012	2152	1237	20345	82%
Mouse	4375	2223	1433	1251	1681	2607	3935	4293	21798	80%
Opossum	4646	2159	2229	2516	2995	3018	2506	1235	21304	78%
Rat	3791	1662	1194	1283	1642	2480	3361	3977	19390	80%
Rabbit	3809	1394	1579	1871	2264	2934	3001	2419	19271	80%
Rhesus	4551	3258	3411	4041	3467	2302	856	0*	21886	79%
Chicken	1369	981	1024	1277	1980	2812	3340	2698	15481	91%

Extended Data Fig. 2 | DDGs. a, Number of DDGs identified in each organ and species using the set of 7,696 1:1 orthologues (left) and the set of all protein-coding genes (right) in each species. The horizontal bar depicts the median. Br, brain; Cr, cerebellum; He, heart; Ki, kidney; Li, liver; Ov, ovary; Te, testis; Te*, testis pre-sexual maturity. **b**, Number of DDGs per species, including number of organs where they show dynamic expression. Asterisk denotes that ovary development is not covered in rhesus macaque, hence there are only 6 organs in total. **c**, Relationship between the number of organs in which genes show dynamic expression and the tolerance to functional variants as measured by: pLI score, RVIS and s_{het} score ($n = 13{,}160$ genes; two-sided Wilcoxon rank-sum test). **d**, Relationship between the number of organs in which genes show dynamic expression and intolerance to duplication and deletion variants (CNV intolerance score; $n = 15{,}728$ genes; two-sided Wilcoxon rank-sum test). **e**, Percentage of organ-specific expressed DDGs at each developmental stage. Bars indicate the range between the replicates. For the brain tissues, DDGs are organ-specific in brain and/or cerebellum. Time points on the x axis are as described in Fig. 1a. **f**, Percentage of transcription factors (TFs) expressed at each developmental stage. Bars indicate the range between the replicates. Time points on the x axis are as described in Fig. 1a. Box plots are as in Fig. 3e.

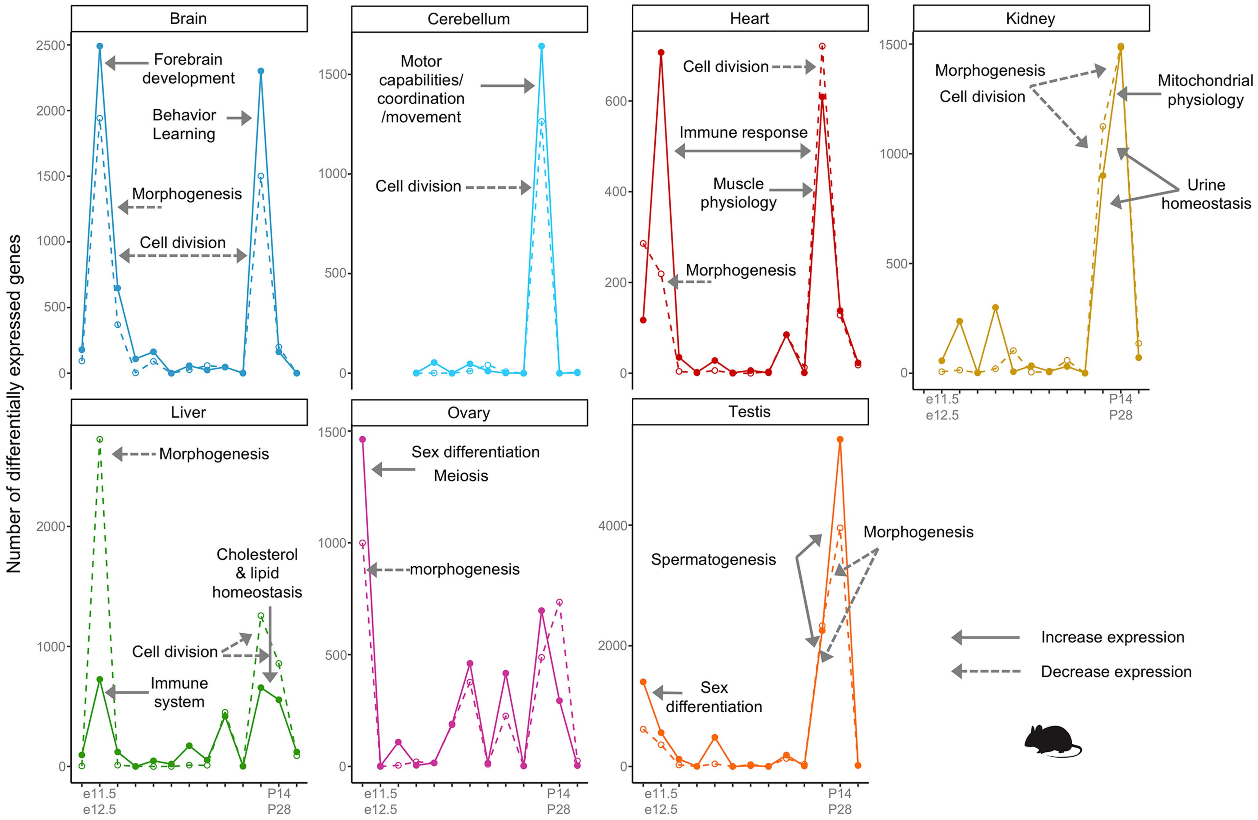
a

	mouse	e10.5	e11.5	e12.5	e13.5	e14.5	e15.5	e16.5	e17.5	e18.5	P0	P3	P14	P28	P63
Hill[12]	human	4wpc	5wpc	6wpc	6wpc	7wpc	8wpc								
de Bakker[13]	human	5wpc	6wpc	7wpc	8wpc										
HuDSeN[14]	human	4-5wpc	6wpc	7wpc	8-8wpc										
Butler[15]	human	4wpc	5wpc	6wpc	7wpc										
This work	human	4wpc	5wpc	7wpc	7wpc	8wpc	11wpc	12wpc	13wpc	19wpc	19wpc	20wpc	toddler	young adult	young middle age
Hill[12]	rat	e12	e13	e14	e15	e16	e17								
This work	rat	e11	e12	e14	e15	e16	e18	e19	e20	e20	P0	P3	P14	P42	P112
This work	opossum	e13.5	P0	P2	P2	P4	P4	P6	P14	P14	P14	P21	P60	P90	P120
Hill[12]	rabbit	e10.5	e12	e13.5	e14.5	e16	e17								
This work	rabbit	e12	e13	e15.5	e16.5	e18	e19.5	e21	e23	e27	e27	P0	P14	P84	P186
This work	chicken							e10	e12	e12	e12	e14	P0	P70	P155
This work	macaque										e108	e130	P23	6 months	9 years

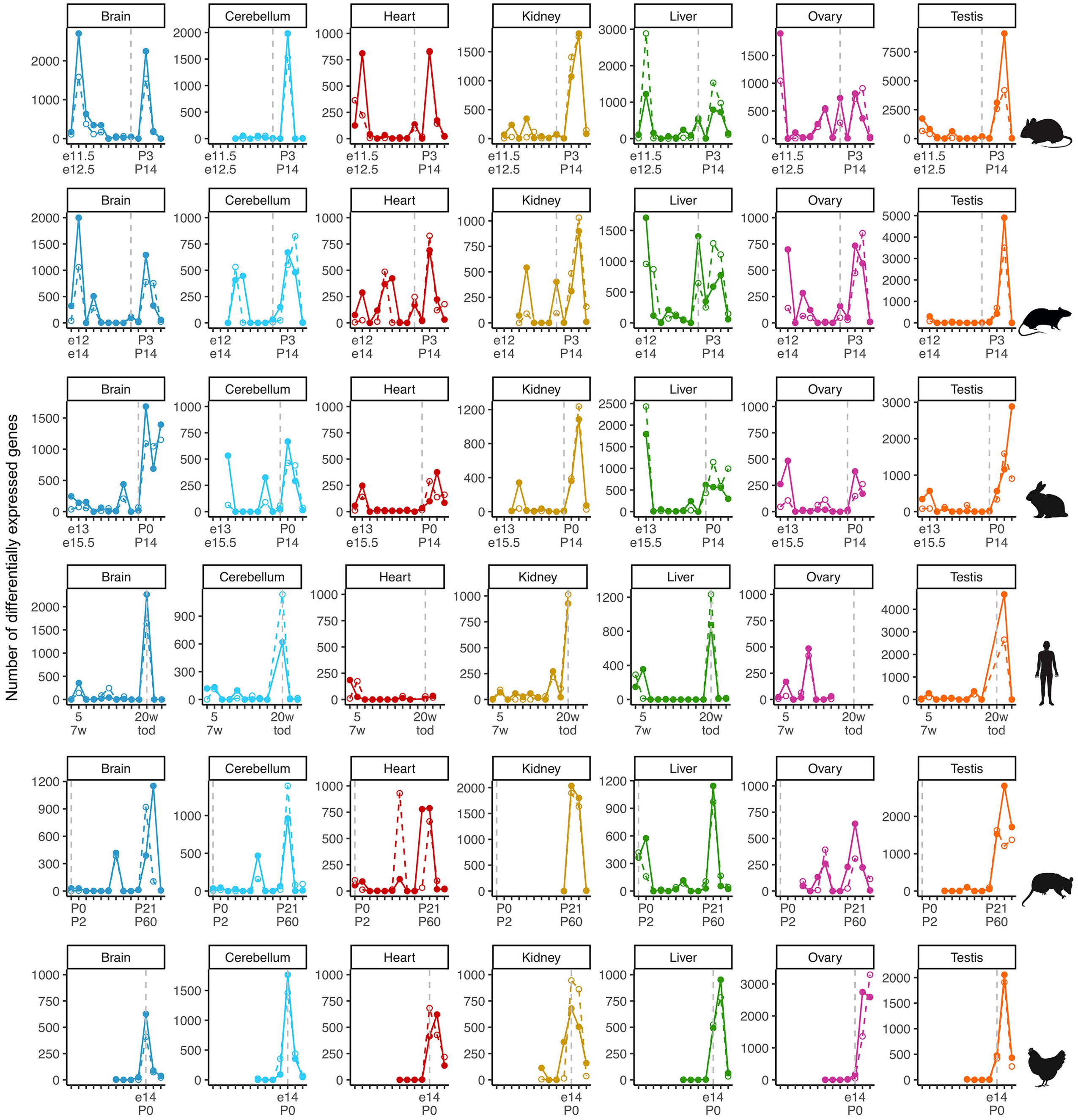
b



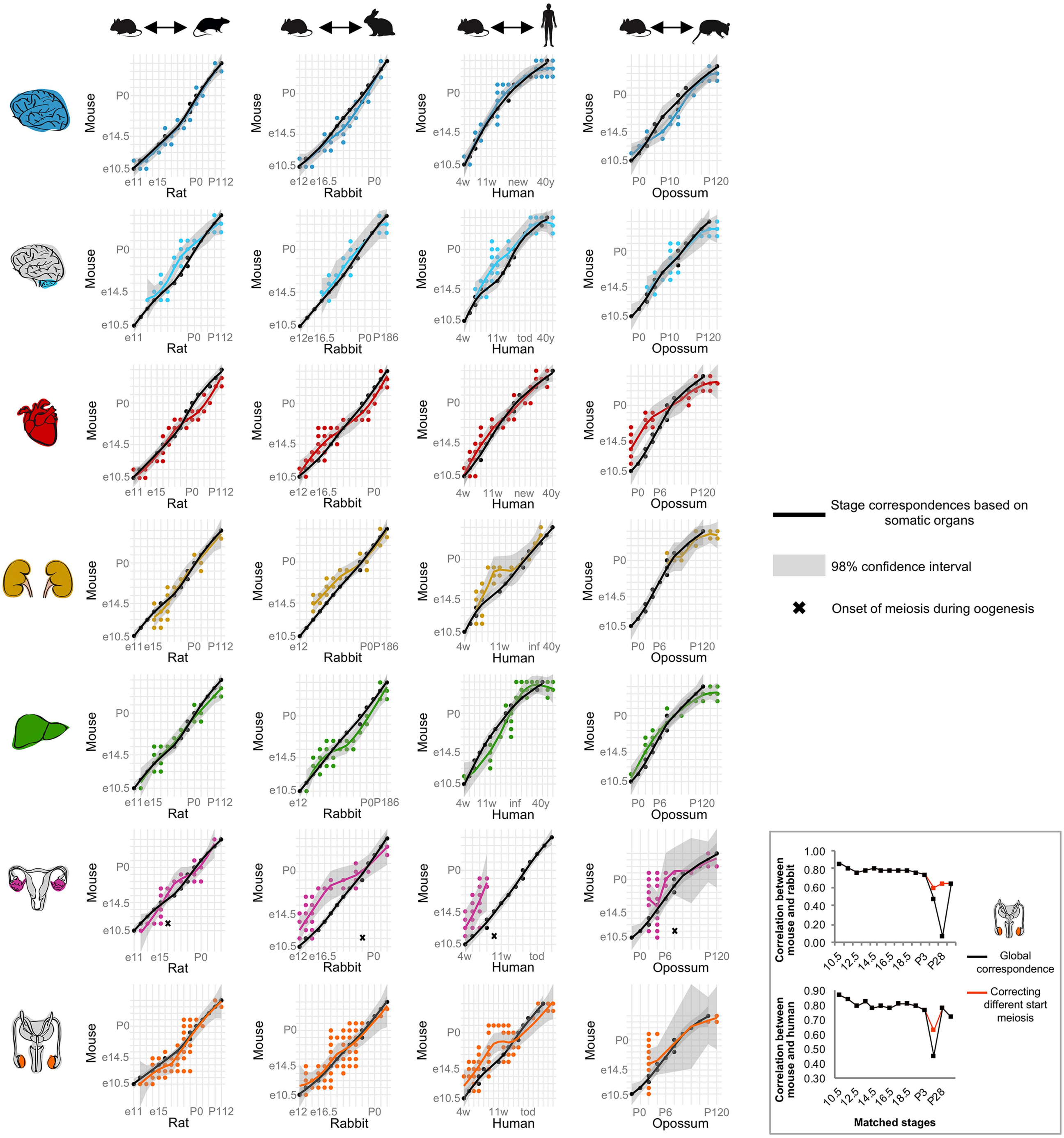
Extended Data Fig. 3 | Developmental correspondences across species. **a**, Developmental stage correspondences established in this study and correspondences based on the Carnegie staging (when available)[12–15]. **b**, Using mouse as a reference, a dynamic time warping algorithm was used to select the best alignment (pink line) between the time series based on stage transcriptome correlations combining all somatic organs ($n = 8{,}940$ genes per organ combinations). New, newborn; tod, toddler (2–4 years); teen, teenager (13–19 years); yma, young middle age (39–41 years); sen, senior (58–63 years).



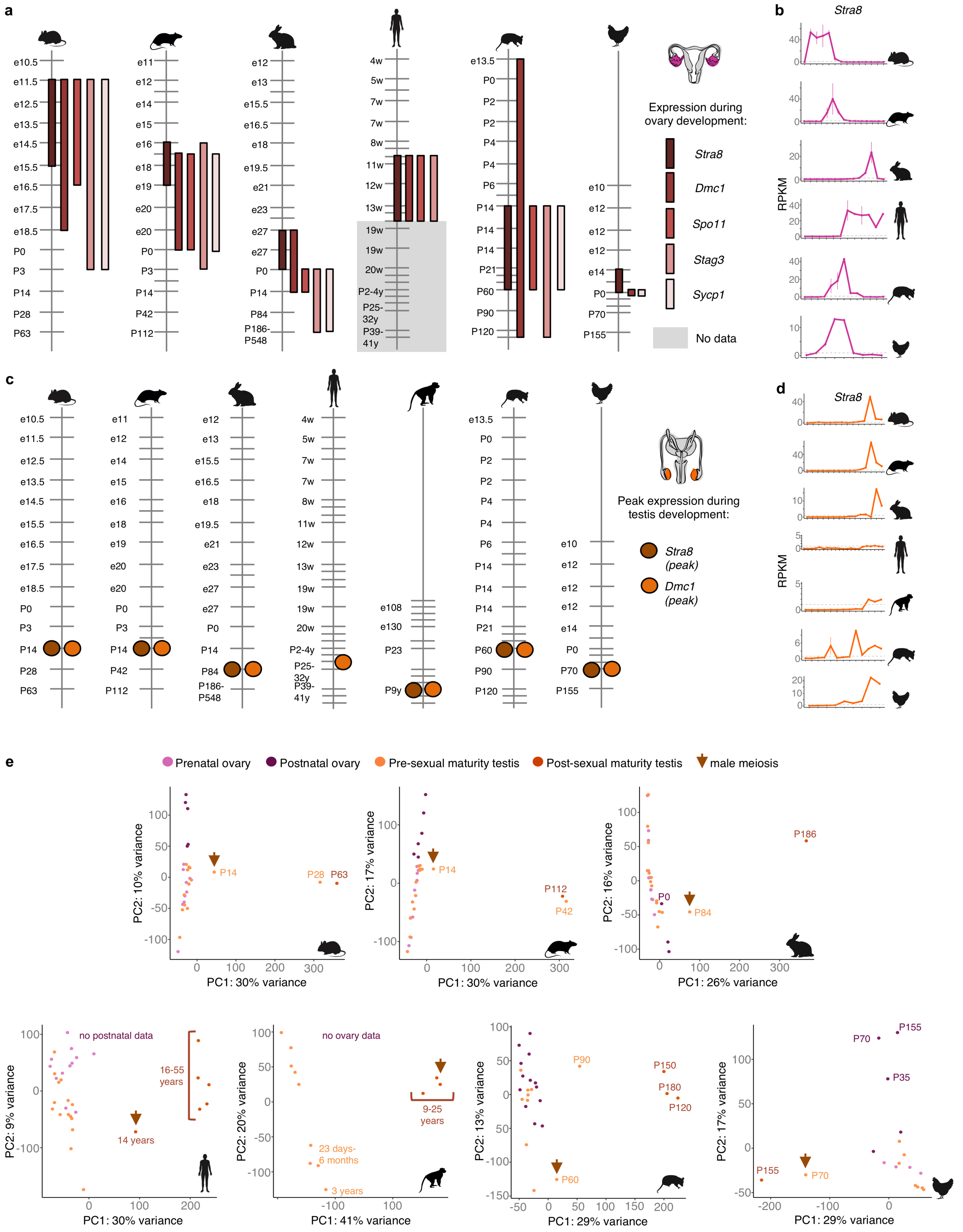
Extended Data Fig. 4 | Periods of greater transcriptional change in mouse. Number of genes differentially expressed between adjacent stages in each organ ($\log_2$ fold change ≥ 0.5). Solid lines refer to genes that increase in expression and dashed lines to genes that decrease. The biological processes and phenotypes enriched at the peaks of differential expression are detailed in Supplementary Table 13.



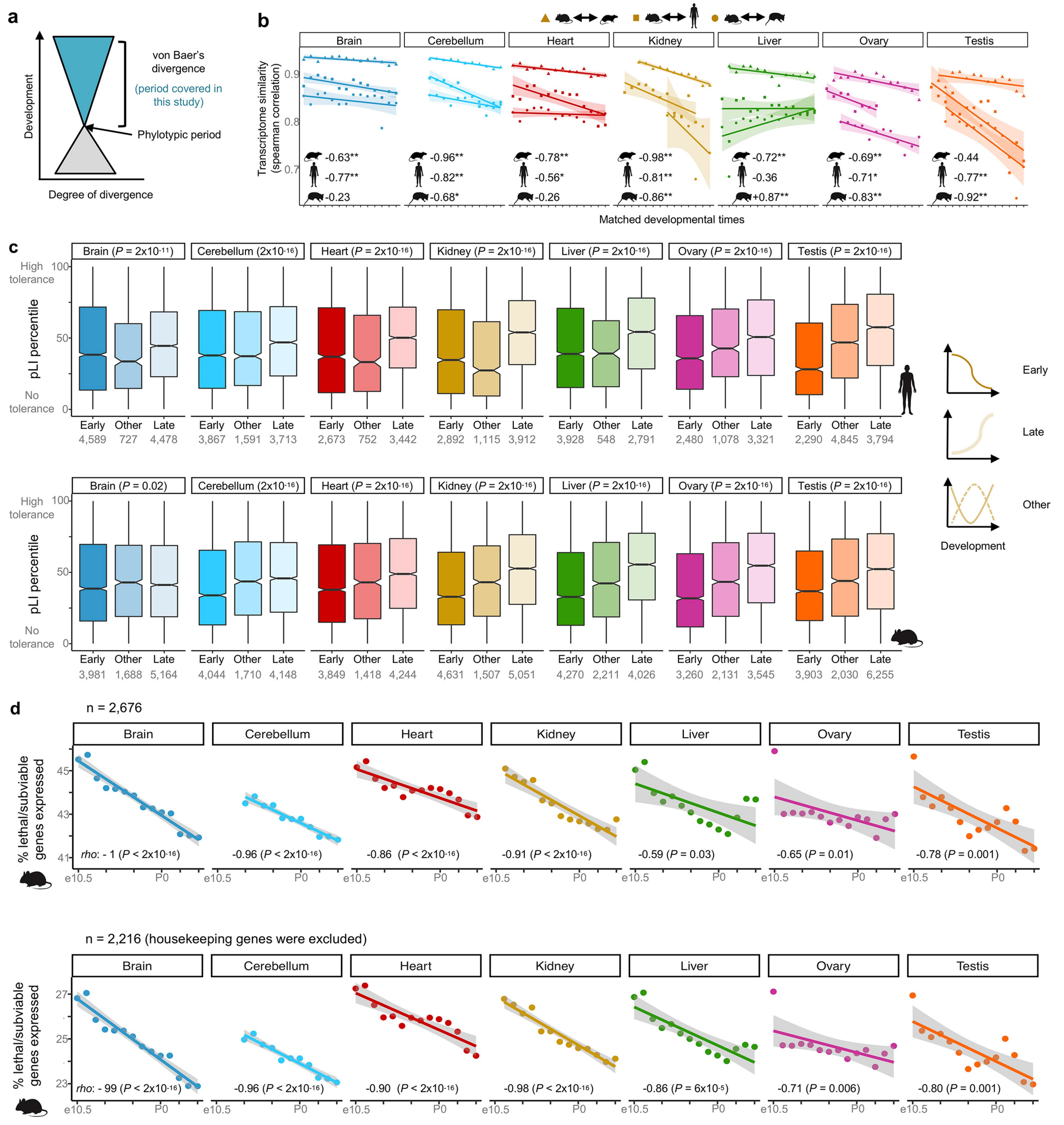
Extended Data Fig. 5 | Periods of greater transcriptional change across species. Number of genes differentially expressed between adjacent, species-matched, stages for each organ ($\log_2$ fold change ≥ 0.5). Solid lines refer to genes that increase in expression and dashed lines to genes that decrease. The vertical dotted line marks birth.



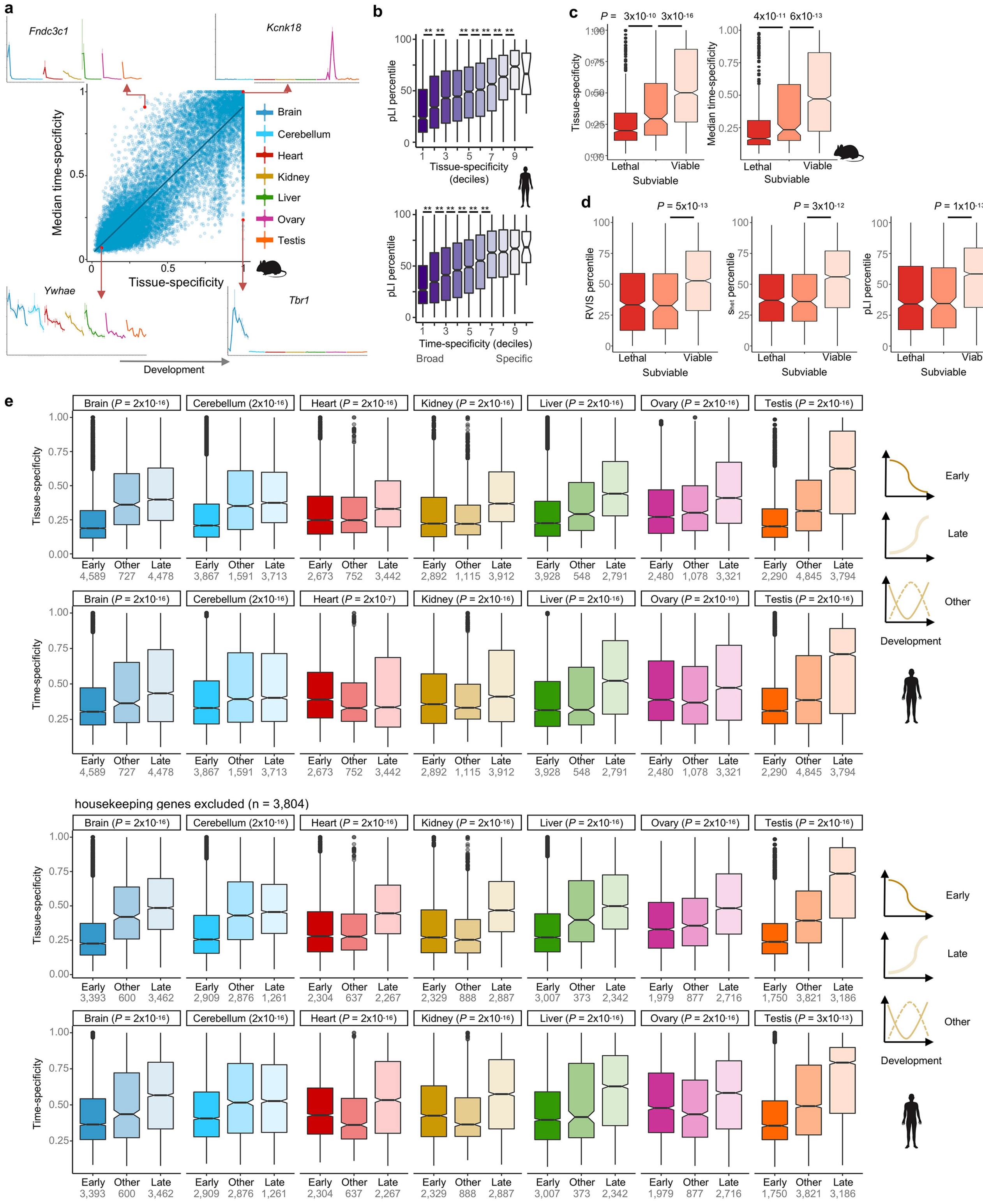
Extended Data Fig. 6 | Organ-specific stage correspondences. Comparison of the global stage correspondences (based on the combined expression of somatic organs; *n* = 8,940 genes per organ combinations; black line) with organ-specific correspondences (based on 2,727 genes for brain, 2,146 for cerebellum, 1,276 for heart, 1,486 for kidney, 1,305 for liver, 1,298 for ovary and 2,153 for testis; coloured lines). With the exception of early heart development in opossum and early ovary development in rabbit and human, the global correspondences are within the 98% confidence interval for predictions computed by the LOESS regression function (local polynomial regression) for each of the organ-specific correspondences (shaded grey area). The same applies to all organs in mouse–chicken and mouse–rhesus macaque comparisons (data not shown). The inset on the bottom right shows the Spearman correlation between mouse and rabbit (top) and mouse and human (bottom) for testis transcriptomes using the global stage correspondences (black line) or adjusting for the different start of meiosis across species (orange line; that is, matching a P14 mouse with a young teenager in human and a P84 rabbit).



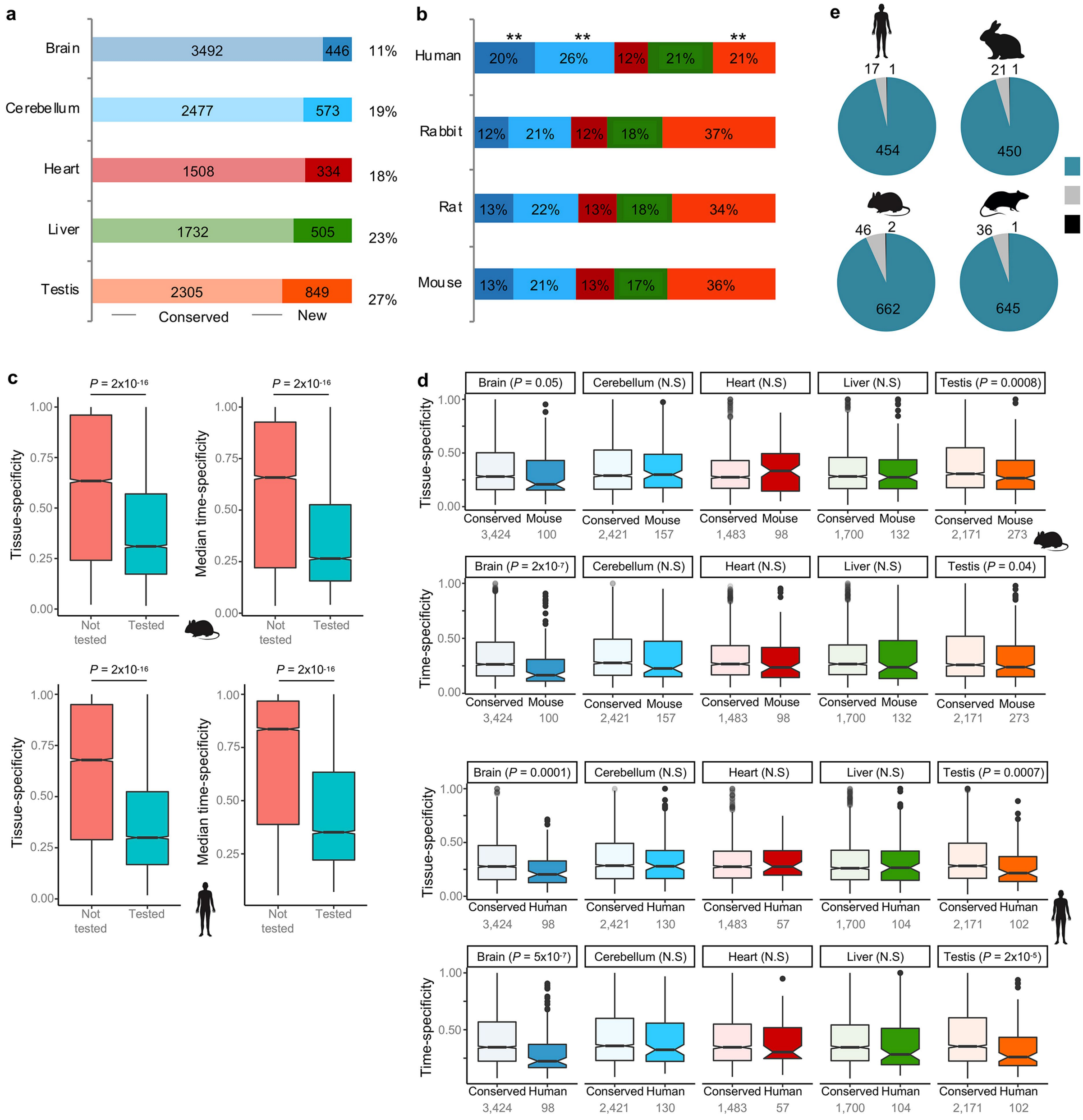
Extended Data Fig. 7 | Heterochronies in gonadal development. **a**, Temporal dynamics of meiotic genes during ovary development. *SYCP1* is not expressed in human ovary. The genes *SPO11* and *STAG3* are not present in the chicken gene annotations used in this work. **b**, Expression of *STRA8* during ovary development. The vertical bars show the range between the replicates and the horizontal dashed line marks 1 RPKM. **c**, Temporal dynamics of meiotic genes during testis development. The profiles of *STRA8* and *DMC1* are represented not by their range of expression but by their highest peak of expression. In rhesus macaque, meiosis is known to start around 3–4 years[36]; our data suggest it had not yet started in the 3-year-old individuals examined. *STRA8* is lowly expressed in the human testis. **d**, Expression of *STRA8* during testis development. The vertical bars show the range between the replicates and the horizontal dashed line marks 1 RPKM. **e**, PCA of ovary and testis development for each species ($n = 21,798$ protein-coding genes in mouse, 19,390 in rat, 19,271 in rabbit, 20,345 in human, 21,886 in rhesus macaque, 21,304 in opossum and 15,481 in chicken).



Extended Data Fig. 8 | Relationships between evolution and development. **a**, Observed relationship between evolution and development. Divergence (horizontal distance) can be morphological or molecular. **b**, Transcriptome similarity between three species pairs throughout development (matched stages) using 11,439 1:1 orthologues. Similar trends were obtained using all species pairs. The weighted average Spearman correlation coefficients are −0.81 ($P = 1 \times 10^{-12}$) for the mouse–rat comparison, −0.69 ($P = 2 \times 10^{-11}$) for mouse–human and −0.42 ($P = 0.0004$) for mouse–opossum. At the bottom are the Spearman correlations between transcriptome correlation coefficients and matched developmental time for each organ and species pair (**$P < 0.02$, *$P < 0.05$). Lines were estimated through linear regression and the 95% confidence interval is shown in the shaded areas. **c**, The pLI score for genes with different developmental trajectories in human (top) and mouse (bottom). Lower values mean less tolerance. The pLI scores used for mouse genes are from their human orthologues. The *P* values refer to early versus late comparisons, two-sided Wilcoxon rank-sum test. Box plots are as in Fig. 3e. **d**, Percentage of lethal and subviable genes expressed throughout development among a set of 2,686 neutrally ascertained mouse knockouts (top) and the same after excluding housekeeping genes (bottom). Spearman correlations at the bottom of each plot. Lines were estimated through linear regression and the 95% confidence interval is shown in grey.



Extended Data Fig. 9 | Pleiotropy as a determinant of the evolution of development. a, Relationship between tissue- and time-specificity. Gene developmental profiles illustrate the extremes of the indexes, which range from 0 (broad time/spatial expression) to 1 (specific time/spatial expression). In the gene plots, the *x* axis shows the samples ordered by stage and organ and the *y* axis shows expression levels. **b**, Functional constraints (measured by pLI score) decrease with increasing time- and tissue-specificity ($n = 9{,}965$ genes). $^{**}P < 0.01$, two-sided Wilcoxon rank-sum test. **c**, Tissue- and time-specificity of mouse genes identified as lethal, subviable, or viable ($n = 2{,}686$; two-sided Wilcoxon rank-sum test). **d**, Levels of functional constraint as measured by RVIS, s_{het} and pLI scores for the human orthologues of genes identified as lethal, subviable and viable in mouse ($n = 2{,}408$; two-sided Wilcoxon rank-sum test). **e**, Tissue- and time-specificity of genes with different developmental trajectories in human (top) and the same after excluding housekeeping genes (bottom). The *P* values refer to early versus late comparisons, two-sided Wilcoxon rank-sum test. Box plots are as in Fig. 3e.



Extended Data Fig. 10 | Evolution of developmental trajectories.
a, Number of genes in each organ that evolved new trajectories across the phylogeny. Includes genes that differ between opossum and eutherians, for which the change cannot be polarized because of the lack of an outgroup. **b**, Distribution of trajectory changes among organs for the different species. The number of genes that changed in each organ is depicted in Fig. 4b. Humans show a relative excess of changes in brain tissues and a relative paucity in testis. $**P = 2 \times 10^{-5}$ for brain, $P = 0.02$ for cerebellum and $P = 1 \times 10^{-10}$ for testis (from binomial tests where the probability of success is derived from what is observed in mouse, rat and rabbit). **c**, Genes tested for trajectory changes (7,020 genes) in mouse (top) and human (bottom) have significantly lower tissue- and time-specificity than genes not tested for trajectory changes (13,325 genes in mouse and 14,778 in human, two-sided Wilcox rank-sum test). **d**, Genes with trajectory changes in mouse (top) and human (bottom) have similar or lower tissue- and time-specificity than genes with conserved trajectories (two-sided Wilcox rank-sum test). N.S., not significant. **e**, Number of organs in which genes evolved new trajectories in the different species. Box plots are as in Fig. 3e.

nature research

Corresponding author(s): Margarida Cardoso-Moreira and Henrik Kaessmann

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	RNA quality was assessed using Fragment Analyzer (Advanced Analytical). Raw reads were obtained using Illumina's HiSeq 2500 control software.
Data analysis	We used GSNAP (22-10-2014) to map the RNA-seq reads against the reference genomes. We used HTSeq (0.6.1) to generate read counts from these alignments. We used EdgeR (3.14.0) to perform normalization and generate the expression tables. We used DESeq2 (1.12.4) to create variance stabilizing transformed counts and to do differential gene expression analyses between adjacent time points. The alignment files were manipulated using samtools (0.1.18) and Bedtools (2.18), and general alignment statistics were created using Picard (1.86). PCAs were done using FactoMineR (1.34). We identified genes with significant temporal changes during organ development using maSigPro (1.44.0). We mapped developmental stages across species using the R package dtw (1.18-1). We identified the most common profiles (clusters) during development using mFuzz (2.32.0). We compared developmental trajectories between species using GPClust. Functional enrichments were done using the R implementation of WebGestalt (0.0.5). All statistical analyses and plots were done in R (3.3.2) as implemented in Rstudio (1.0.136). Plots were created using the R packages ggplot2 (2.2.1), gridExtra (2.2.1), reshape2 (1.4.2), plyr (1.8.4), and factoextra (1.0.4).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw and processed RNA-seq data are available from ArrayExpress: E-MTAB-6769 (chicken), E-MTAB-6782 (rabbit) E-MTAB-6798 (mouse), E-MTAB-6811 (rat), E-MTAB-6813 (rhesus), E-MTAB-6814 (human) and E-MTAB-6833 (opossum). We also created a publicly available data resource (evodevoapp.kaessmannlab.org), where the profiles of individual genes can be easily visualized and the expression tables can be downloaded.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Our aim was to study gene expression profiles for 7 major organs from early organogenesis to adulthood in human, mouse, rat, rabbit, rhesus macaque, opossum and chicken. We started sampling as early as it was possible to dissect and isolate the organs and then sampled prenatal development at regular intervals. Postnatally we sampled around the major developmental milestones. The time points sampled were chosen to cover the most important periods of organ development and varied from 9 time points in chicken to 23 in human. We aimed for 4 biological replicates (2 males and 2 females) for somatic organs and 2 replicates for the gonads. Fewer replicates were available for the primates. The number of replicates were chosen based on field standards for differential gene expression analysis; no statistical methods were used to pre-determine sample size. This resulted in 1,893 samples (full details in Supplementary Tables 1 and 2). This dataset provides a detailed, quantitative description of gene expression throughout the development of the cerebrum, cerebellum, heart, kidney, liver, ovary and testis for 6 mammals and a bird.
Data exclusions	We are making available a small number of libraries that were not used in this study because they had a correlation with their biological replicates lower than 0.90. The decision to exclude these libraries was made before performing the analyses described in the manuscript. We are providing these libraries because they were used in other projects from our group (still to be published) and are therefore part of the evo-devo resource. These libraries are clearly marked in Supplementary Table 2.
Replication	We generated biological replicates for the stages and organs sampled in all species. We aimed for 4 biological replicates (2 males and 2 females) for somatic organs (2 for primates) and 2 replicates for the gonads. The analyses described in the manuscript take into consideration the information from the biological replicates. We used PCA and hierarchical clustering to identify and exclude outlier libraries (e.g caused by low RNA quality). We also excluded libraries that showed a Spearman's correlation coefficient with its biological replicates lower than 0.9.
Randomization	All comparisons in this work are based on 3 biological variables: species, organ and developmental stage. Generally, randomization does not apply. When sequencing the RNA-seq libraries (they were multiplexed in sets of 6 or 8) we mixed samples from different organs, stages and species. Full randomization was not possible because the samples arrived at different times and were processed based on their date of reception.
Blinding	Blinding was not relevant to our study. Both data collection and analyses required an understanding of the nature of the sample being collected/analyzed (i.e. species, organ, developmental stage).

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	All mouse (Mus musculus) samples are from the strain CD-1 (RjOrl:SWISS); all rat (Rattus norvegicus) samples are from the outbred strain Holtzman SD; all rabbit (Oryctolagus cuniculus) samples are from the outbred New Zealand breed; all chicken (Gallus gallus) samples are from the red junglefowl, the progenitor of domestic chicken. The other species used in this study were the gray short-tailed opossum (Monodelphis domestica) and rhesus macaque (Macaca mulatta). We sampled males and females in each species.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.

Human research participants

Policy information about studies involving human research participants

Population characteristics	We sampled organs (i.e. forebrain/cerebrum, hindbrain/cerebellum, heart, kidney, liver, ovary and testis) from both males and females, starting at 4 weeks post conception and ending at 60 years of age. The organ, developmental stage and sex of each sample is described in Supplementary Table 2.
Recruitment	There was no direct recruitment for this work. The human prenatal samples were provided by the MRC-Wellcome Trust Human Developmental Biology Resource (HDBR) and were derived from elective abortions with normal karyotypes. The tissue donations were made entirely voluntarily by women undergoing termination of pregnancy. Donors were asked to give explicit written consent for the fetal material to be collected, and only after they had been counselled about the termination of their pregnancy. The human postnatal samples were retrieved from the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland (USA) and from the Chinese Brain Bank Center (CBBC) in Wuhan (China). They originated from individuals with diverse causes of death that, given the information available, was not associated with the organ sampled. Written consent for the use of human tissues for research was obtained from all donors or their next of kin by the respective tissue banks. We are not aware of any potential self-selection biases (or other) that could have affected this work.