Restructured genomes may have shaped the evolution of clade-specific traits in mammals

Jui Bhattacharya, Mohan Lal, Kuljeet Singh Sandhu*

Department of Biological Sciences
Indian Institute of Science Education and Research (IISER)- Mohali
Knowledge City, Sector 81, SAS Nagar 140306, Punjab, India.

*sandhuks@iisermohali.ac.in

Keywords: Genomic rearrangements, mammalian evolution, clade-specific traits

Abstract
Genomic rearrangements (GRs) can drive the evolution of genome order. While their role in speciation is well established, the underlying adaptive significance remained underappreciated. Through analysis of 71 mammalian genomes, we inferred the clade-specific evolutionary loss of synteny through GRs. The genes near the clade-specific break-point regions were mostly associated with the clade-specific traits and exhibited greater rates of expression divergence in the clade that had lost the synteny. The disruption of otherwise syntenic chromatin contacts and the exchange of chromatin domains with contrasting epigenetic states explain the underlying mechanism of expression divergence. The observations imply that the clade-specific genomic rearrangements likely guided the evolution of clade-specific traits and may have wide-spread adaptive significance.

Introduction

Genomes have gone through several rounds of structural rearrangements during the course of evolution. These rearrangements alter the genome order and, therefore, may change the regulatory context of the genes. Such alterations may lead to expression divergence, of genes popularly known as chromosomal position effect[1,2]. Ectopic position effects can lead to diseases in human and mice[3,4]. For example, translocation breakpoint between PAX6 and its downstream regulatory elements associates with Aniridia in human[5]. A genomic rearrangement at 100kb upstream to Steel(Sl) locus downregulates Mgf gene and causes sterility in female mice[6] . Similarly breakpoints near SOX9, TWIST, and RIEG genes have association with campomelic dysplasia, Saethre-Chotzen, and Rieger syndromes respectively in human[7–9]. Yet the evolutionary implications of chromosomal position effects remain largely under appreciated. De et al had shown that expression divergence associated with the position effects may contribute to phenotypic differences between human and chimpanzee10. Bagadia et al showed that the genomic rearrangements may explain the evolutionary divergence of some brain related traits in rodents11. Recently, Real et al12 identified genomic rearrangements that altered the regulatory context of gonadal gene expression governing intersexuality in moles. These studies were limited by the number of species analysed and by the context of genomic rearrangements. With the availability of large number of mammalian genomes, it is now possible to examine the phenotypic associations of genomic rearrangements in mammals. Through analysis of 71 mammalian genomes representing 8 clades, we identified 206 clade-specific long-range genomic rearrangements marking evolutionary loss of synteny. We found that the clade-specific breakpoints flanked by the genes associated with specific phenotypes of respective clades. The clade-specific breakpoints frequently disrupted the conserved enhancer-promoter communications, which coincided with the diverged expression rates of the flanking genes. Our observations implied the wide-spread adaptive significance of genomic rearrangements in mammals.

Results

By applying the strategy elaborated in the Methods section, we identified the long range (>10Mb and *trans*) genomic interruptions of neighbouring domains. In this study, we focused upon the instances of inferred loss of synteny. We observed that many breakpoints nonrandomly cluster in genomic domains of around 1.4Mb size. Therefore, to remove the regional bias, we clustered together the breakpoints constrained within 1.4 Mb and termed these as evolutionary break-point regions (EBRs) (Figure S1). The inferred loss of synteny was also confirmed by analysing the inferred chromosome structures of eutherian ancestor directly obtained from Kim et al[13]. Around 90% of the gene pairs were <=3 genes apart, 7% were >3 gene apart and 3% were located in *trans* in the eutherian ancestor, largely validating our inference of ancestrally syntenic configuration of identified cases (Figure S2). In total, we identified 206 instances of clade-specific loss of synteny (cs-LoS) across 71 mammals. The artiodactyls and rodents exhibited the most number of cs-LoS events (38 and 94 respectively), which was coherent with the earlier reports[13,14].

To assess if the cs-LoS EBRs position preferably near certain functional loci, we obtained the genes within and the flanking 10 genes on the either sides of cs-LoS EBRs. Gene ontology analysis yielded enrichment of clade-specific functional terms. Primates exhibited enrichment of glucocorticoid secretion, (CRHR1, TAC1, CRY1

etc.), lipid transport, synaptic transmission and behaviour related terms (SHANK2, RELN, FOXP1, GIT1, KLHL1, NRG1, MAPT etc.) (Figure 2). These functions have associations with primate-specific trait divergence[15–18]. The genes flanking artiodactyl-specific LoS break-points were enriched with flavonoid and lipid metabolism related terms, including genes like UDP-glucuronosyltransferase 1A (UGT1A) and Fatty Acid Binding Protein 4 (FABP4) (Figure 2). These terms associate with vegetarian diet and musculoskeletal health respectively, which coincide with herbivory and relatively large body mass of artiodactyls[19,20]. Carnivore-specific LoS exhibited enrichment of extra-cellular matrix proteins, particularly matrix metalloproteases (MMPs) (Figure 2). It is established that carnivores, unlike other mammals, have endotheliochorial placenta[21]. Differential expression of MMPs is the hallmark of invasive placentation in carnivores[22]. We further hypothesize that enrichment and expression divergence of MMPs in carnivores may also relate to effective wound healing in carnivores being susceptible to frequent wounds due to predation and the territory defence. Rodents exhibited enrichment of renal filtration related terms, aligning to their higher glomerular filtration rate, which is an adaptation to physiological requirements in diverse habitats and dietary habits of rodents[23–25]. The term includes the genes Anillin, Glucocorticoid Induced 1 (GLCCI1), Thrombospondin type-1 domain-containing 7A (THSD7A), Apolipoprotein L1 (APOL1) etc., which have strong associations with renal filtration and implicate in kidney diseases[26–29]. Top enriched genes in Chiroptera-specific LoS break-points exhibited immunity related functions, adhering to pronounced immune tolerance in bats[30,31]. Among others, we also observed certain genes, like CEBPA and TNFRSF1A, that are associated with echolocation in bats[32,33]. Afrotheria-specific LoS showed enrichment of potassium ion channels, vitamin-K metabolism, and the blood coagulation (Figure 2). Ion transporters are linked with the cell proliferation, cell size and cancer, the traits significantly diverged in afrotheria clade[34]. Vitamin-K is linked with the osteoblast differentiation, likely associated with the larger bodymass of afrotherians like elephant, manatee etc. Blood coagulates faster in elephants[35] and the presence of blood coagulation factors in our set might mark an adaptation against an ancient fatal haemorrhagic disease caused by elephant endotheliotropic herpesvirus[36]. Similarly, perissodactyl-specific breakpoints showed moderate enrichment of aerobic respiration and the cetacean break-points shows some enrichment of r-RNA processing and sterol related functions, aligning to remarkably high aerobic respiration capacity of equines[37] and longevity, bodymass and diverged metabolism of cetaceans[38,39] (Figure 2).

In most clades, genes proximal to the cs-LoS breakpoints exhibited the expression divergence from that of the phylogenetically proximal clade controlled through an outgroup, implying the adaptive significance of genomic rearrangements (Figure 3a). Chromosomal position effect may explain the observed expression divergence. To test this, we first analysed the chromosomal coordinates of enhancer-promoter pairs that exhibited evolutionarily conserved synteny. We observed that the synteny of enhancer-promoter pairs was lost in the species exhibiting cs-LoS events when compared with the null distribution, suggesting that the expression divergence of genes might have been driven by the loss of proximity to the cognate enhancers (Figure 3b-c). We further assessed the position effect through analysis of chromatin compartments. We calculated Kronecker's delta (δ_{AB}) for the consecutive windows along ±1Mb regions around the cs-LoS breakpoints. Drop in average δ_{AB} at the breakpoints highlighted change in compartments. We observed that the

compartments with the contrasting states (A and B) in a species tend to fuse together and subsequently exhibit a coherent state (A or B) in the species exhibiting the respective cs-LoS event (Figure 3d-e). This implies the alteration in gene expression through mixing of contrasting epigenetic states *via* gained proximity. This is illustrated through an example of primate-specific genomic rearrangement near Neurokinin-1 (TAC1), a glucocorticoid secretion and behaviour related gene. The gene locates in B-compartment in mouse, which gains proximity to an A-compartment, and subsequently switches from B-to-A compartment in neuronal cells (Figure 3e). Overall the split of enhancer-promoter synteny and gain of proximity between A and B chromatin compartments highlight the mechanistic basis of expression divergence of genes located near cs-LoS breakpoints.

We further exemplify a primate-specific LoS event near Corticotropin Releasing Hormone Receptor 1 (CRHR1) gene that coordinates the autonomic, endocrine and behavioural response to stress and depression through Hypothalamus-Pituitary-Adrenal (HPA) axis[40] and may also regulate the timing of birth[41]. Interestingly, CRHR1 shows primate-specific conservation of 29 amino acids of exon-6, and insertions of retroviral LINE elements at the upstream region (Figure S3). The gene also exhibits loss of chromatin contacts and expression divergence in primates (Figure 4). We propose that the sequence and the expression divergence of this locus may link to primates' adaptation to stress from complex social structures[42,43]. Artiodactyl-specific LoS near UGT1A genes, which have expanded copy numbers in harbivores[20], exhibited loss of chromatin contacts and expression divergence in artiodactyls (Figure 4). We report similar examples for other clades in the Figure S4.

Discussion

Genes are not autonomous transcriptional units within the genomic landscape but rather operate in a highly context-dependent manner influenced by their local and global chromatin environment. Reporter integration in proximity to enhancer strengthens the expression, while that within lamina-associated domains attenuates the transcriptional activity[44]. Distinct linear and spatial positions of a gene also confer variation to its expression noise[45,46]. The alterations in genomic context through chromosomal rearrangements can, therefore, influence the epigenetic and transcriptional states of the genes. Although studied in somatic human cells and cancer tissues[47,48], the evolutionary dynamics of such position effects remains poorly understood. Our observations through this study highlight that the clade-specific long-range genomic rearrangements associate with the enrichment and expression-divergence of genes with clade-specific traits. Interestingly, these genes have also lost the proximity to their cognate enhancers as reflected from the splitting of enhancer-promoter pairs, likely representing one of the possible causes of the observed divergence of gene expression in the respective clades. The gained proximity of chromatin domains with contrasting epigenetic states further strengthens the notion of chromosomal position effect in evolutionary context. We propose that the alteration in regulatory context may have contributed to the expression evolution of genes proximal to clade-specific EBRs. From our observations, it follows that clade-specific GRs may have guided the evolution of clade-specific mammalian phenotypes through expression evolution. While largely correlative in nature, our observation may guide future experiments and design genetic manipulations in closely related species to test the causal associations.

The reported genomic rearrangements and the engaged loci also offer a valuable resource to prioritize biomedical research. Many of the reported loci also implicate in human diseases and understanding the rewiring of their regulatory contexts could help in genetic prioritization and devising medical strategies to intervene the mis-regulation. Certain mammals are resistant to certain diseases, like elephant is resistant to cancer and bats are immune tolerant to viruses. Our observations may also help studying the regulatory contexts which are instrumental in disease resistance and exploit that knowledge to understand disease susceptibilities in humans.

Materials and Methods

Genome data

We obtained the chromosomal coordinates of human orthologous genes for 18 mammalian species from Ensembl (http://www.ensembl.org). Fifty three genomes not represented in Ensemble were obtained from DNA Zoo (https://www.dnazoo.org/assemblies) and Bat1K Project (https://bat1k.com/). We aligned the genome FASTA files to the human genome using LASTAL(https://github.com/lpryszcz/last). Resulting maf files were processed with *'last-split'* and converted to chain files using *'maf-convert'* utility (https://github.com/lpryszcz/last). To obtain orthologous genomic coordinates, we used the UCSC LiftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver).

Identification of clade-specific loss-of-synteny (LoS) events

To define evolutionary breakpoints of long-range genomic rearrangements, we used a distance-based approach. For each species, we first calculated linear distances between non-redundant all possible *cis* gene pairs. Gene pairs separated by ≤1 Mb were considered syntenic in a given species. The corresponding intergenic distances between orthologues of these syntenic pairs were then obtained for all other species and their distribution was visualized as histograms using the 'hist', function in R. The analysis revealed a bimodal distribution with two modes separated at around 10 Mb distance. This 10 Mb threshold was, therefore, used as a threshold to define loss of synteny events. Gene pairs that were within 1 mb distance in one species but exceeded 10Mb in other species were categorized as *cis*-rearrangement. If the orthologues of syntenic pairs were located on different chromosomes in another species, they were classified as *trans*-rearrangements. To minimize noise that might result from retrotransposition or assembly errors, we further obtained the set where the orthologous disruption occurred with at least 3 genes on either side of the break-point. We further calculated the inter-breakpoint distances defined as the linear distances between gene pairs forming breakpoints across all species. The distribution of distances suggested a distinct natural cut-off of 1.4 Mb, below which the breakpoints can be considered as clustered into evolutionary break-point region (EBRs). An EBR was identified as a clade-specific loss-of-synteny (LoS) event if it engages in LoS events in at least 50% of the species within a given mammalian clade, while being syntenic in most other clades (split in not more than two species).

Gene Ontology analysis

For each evolutionary breakpoint region (EBR), we extracted all genes within EBR and 10 genes upstream and downstream of both ends. These unique gene sets were compiled for each clade. Gene ontology (GO) enrichment analysis was performed using the ToppFun function from the ToppGene Suite (https://toppgene.cchmc.org/) to fetch the GO Biological Process categories. P-values were corrected using Benjamini-Hoschberg method.

Gene expression analysis

We retrieved the expression data of CRHR1 and UGT1A1 of 35 human tissues from the TissueEnrich database(https://tissueenrich.gdcb.iastate.edu/) for tissue specific gene enrichment analysis. To calculate expression divergence, we used tissue-wide expression data from the Bgee database(https://www.bgee.org/). For each gene (CRHR1 and UGT1A1), a divergence ratio was calculated as the median of across-clade versus within-clade distances. These gene-specific ratios were compared to the genome-wide median divergence ratios calculated across all genes.

Expression rate analysis

We used 'TreeExp' R-package(https://github.com/hr1912/TreeExp) to evaluate the evolutionary rates of expression divergence assuming Ornstein-Uhlenbeck model of gene expression evolution. For each clade, we assigned a test species, a reference species and an out-group species from phylogenetically adjacent clade. We made three distance-based bins, each comprised of 25 genes, on either side of the break-point. Gene expression data were obtained from Liu et al[49], which provides expression profiles for three tissues: brain, kidney, and liver for multiple species.

Enhancer-promoter contact analysis

All human enhancer-promoter (EP) pairs and their orthologues in other species, were obtained from the dataset provided by Laverré et al[50]. The evolutionarily conserved EP pairs were then used to analyze disruption by clade-specific LoS events. If a conserved EP pair flanked a breakpoint unique to a clade, the pair was considered split by the rearrangement. For clades such as Chiroptera, Perissodactyla, Afrotheria, and the suborder Hystricomorpha, for which the enhancer orthologs were not included in the dataset by Laverré et al, the coordinates were lifted over using the UCSC liftOver tool (https://genome-store.ucsc.edu/). To calculate the statistical significance of observed EP pair disruptions, we performed bootstrap by randomizations breakpoint coordinates across the human genome 1000 times while keeping the chromosomal distribution of breakpoints same as original set. The bootstrap-based p-values were obtained using:

$$p = \frac{1}{B}\left(1 + \sum_{b=1}^{B} (1)_{k_b > k}\right)$$

where B is the total number of bootstrap samples, k is the observed number of disrupted EP pairs.

Hi-C analysis

We used HUGIn web browser (https://yunliweb.its.unc.edu/hugin/tutorial_page.php), to fetch virtual 4C interaction profiles at 40 kb resolution from human Hi-C data across relevant tissues. For mouse (mm10), .cool files were obtained from the 4D

Nucleome Data Portal (https://data.4dnucleome.org/) and virtual 4C profiles were generated using Fit-Hi-C (https://github.com/ay-lab/fithic). We downloaded chromatin compartment (A/B) files for both human and mouse tissues from the 4D Nucleome. Change in compartment along the genomic regions spanning the break-points was calculated using Kronecker's delta function. If consecutive genomic bin i and bin $i+1$ had same compartment (A-A or B-B), the function takes the value 1, and zero when i and $i+1$ bin had different compartment values (A-B or B-A). We calculated the average and 95% confidence intervals of Kronecker's delta across all breakpoints for each tissue and clade (supplementary). For compact representation a global average of Kronecker's delta was plotted at 250 kb resolution around +/- 1 Mb to the break-point.

Acknowledgement
Authors duly acknowledge the financial support from Department of Biotechnology, India (BT/PR40149/BTIS/137/36/2022, BT/PR40198/BTIS/137/56/2023).

Figure legends

Figure 1. Inference of clade-specific loss of synteny events across mammalian lineages. **(a)** Schematic representation of the strategy to define the clade-specific loss of synteny events. The distribution plot of genomic distance between adjacent genes in a species, which were <1 Mb apart in another species. Vertical line demarcates the separation of gene-pairs by >10mb in a species. **(b)** Heatmap representation of number of genomic rearrangements between lineages **(c)** Inferred clade-specific loss of synteny (LoS) events. Highlighted in rectangles are the clade-specific LoS events.

Figure 2. Enrichment of GO Biological Process terms among genes flanking the clade-specific EBRs. We implemented Benjamini-Hoeschberg correction of multiple testing. The vertical lines mark the FDRs of 0.05.

Figure 3. Expression divergence and position effect associated with the clade-specific loss of synteny. (a) Lineage-specific expression divergence of the genes as a function of distance from the clade-specific evolutionary break-points, calculated assuming Ornstein-Uhlenbeck model of transcriptome evolution. Each box represents the distribution of expression divergence values of 25 genes incrementally distant from the breakpoint. A species from the nearest order or sub-order was taken as reference species, while a species hierarchically outside the clade of test and the reference species was taken as outgroup species. (b) Enrichment of clade-specific breakpoints within and around enhancer-promoter loops in the species having syntenic configuration. Loop coordinates were extended either side by half of the loop length-L (b) The number of conserved enhancer-promoter loops that had lost proximity through GR in a clade-specific manner, superimposed onto the random null distribution obtained from controlled randomized of breakpoints. P-values are calculated using bootstrap method. Shown are the plots for the cumulated set and the clade-specific sets. (d) Kronecker's δ_{AB}, signifying change in A and B compartment states of neighbouring domains, as a function of distance from the break-points in primates and rodents. Drop in Kronecker's delta signifies the switch in states (A-to-B or B-to-A). The three plots correspond to syntenic configuration in mouse (left), distal regions (syntenic in human) when juxtaposed in

mouse (middle), rearranged syntenic regions in human (right) respectively in case of primate-specific LoS. Similarly, for rodent-specific LoS, syntenic configuration in human (left), distal juxtaposed regions in human (middle) and rearranged syntenic regions in mouse (right) are plotted. Chromatin compartment states (leading Eigen vector of Hi-C contact matrix) for regions around TAC1 locus, which exhibited primate-specific LoS nearby, in mouse and human neuronal cells. A and B compartments are coloured as red and grey colours respectively.

Figure 4. Illustrative examples of clade-specific breakpoints.(a-b) Left-panels: CRHR1 and UGT1A expression values in 35 human tissues respectively. Right-panels: expression divergence calculated as the Euclidean distance across tissues and summarized as ratio of across-clade to within-clade distances.(b-c) Mouse-to-human and human-to-cow synteny plots around (b) CRHR1, and (c) UGT1A gene respectively. Shown below are the tracks of RefSeq non-redundant genes, H3K27ac modification, and virtual 4C in relevant tissues. The clade-specific break-point regions are highlighted by vertical bars. 4C reference loci are highlighted using arrow-heads. Black and red curves represent observed and expected 4Csignals respectively. Green curve represents -log10 of p-values. Blue dashed line represents the p-value cut-off of 0.05.

Figure 5. Graphical summary of the observations.

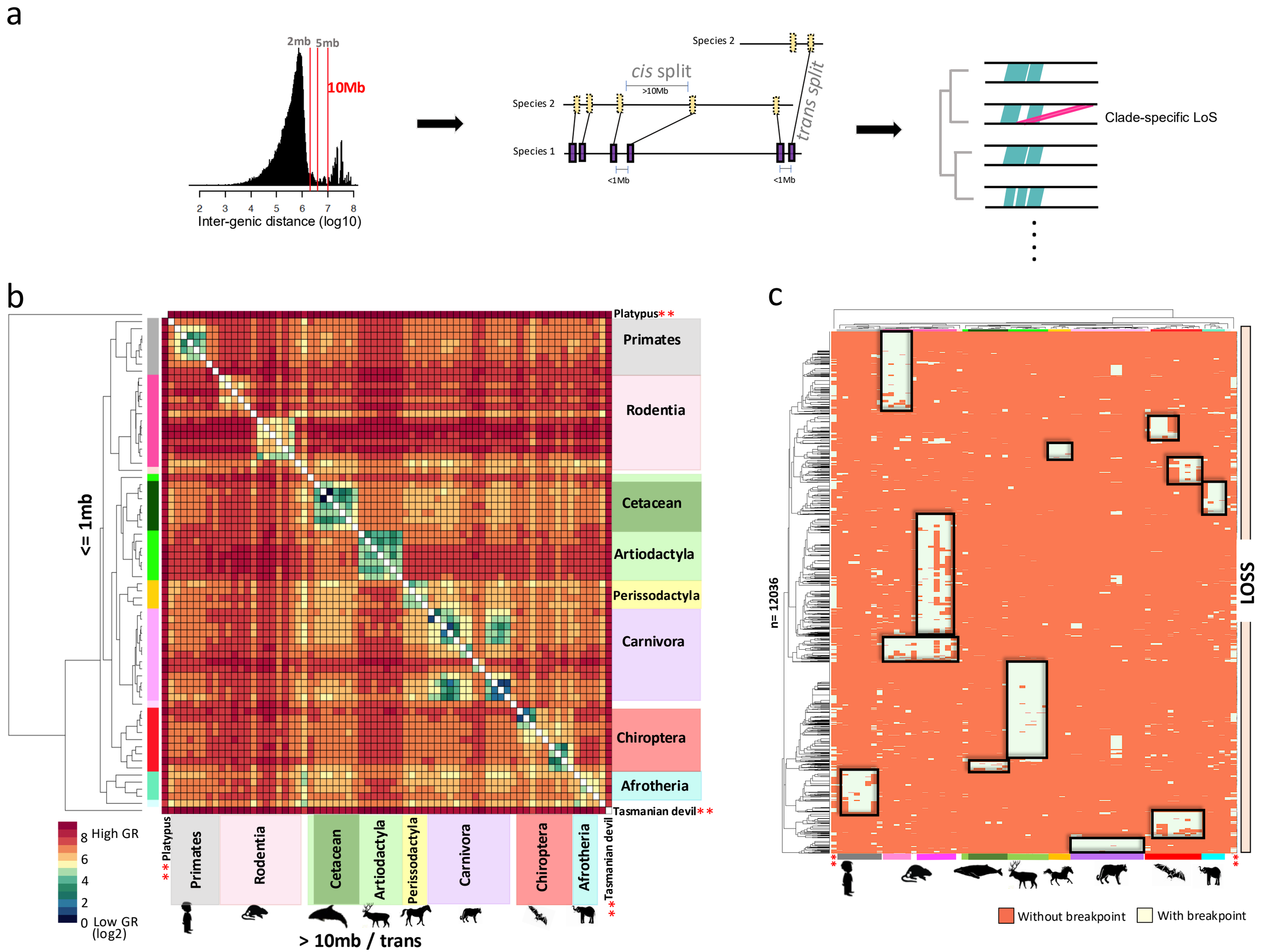
References:

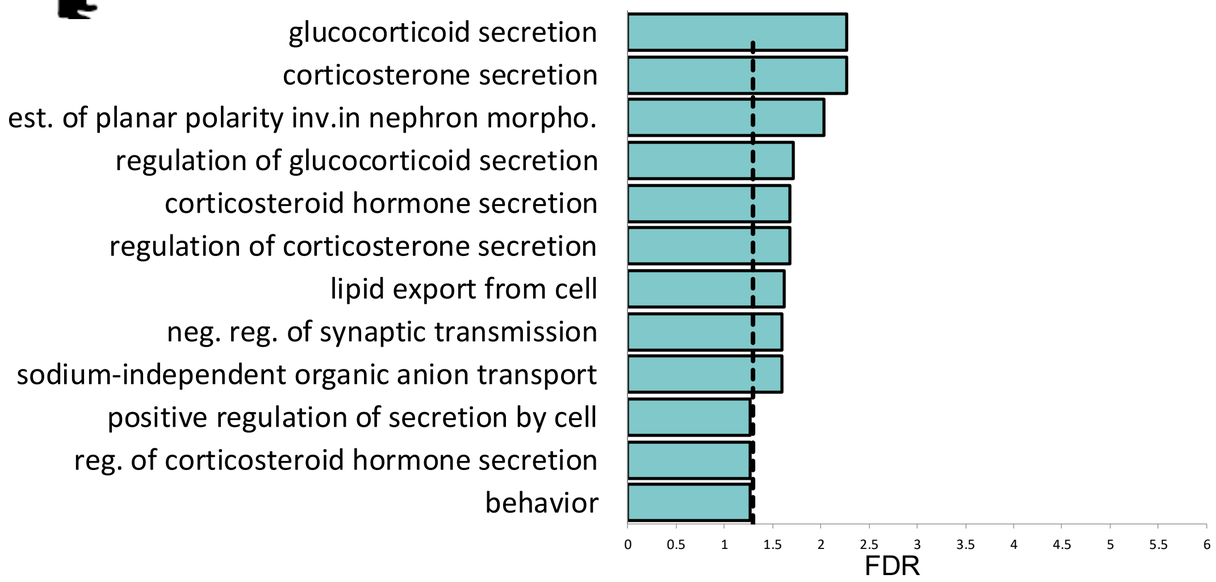
1. Elgin, S.C., and Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. Cold Spring Harb Perspect Biol *5*, a017780.
2. Wallrath, L.L., and Elgin, S.C. (1995). Position effect variegation in Drosophila is associated with an altered chromatin structure. Genes Dev *9*, 1263–1277.
3. Kleinjan, D.J., and Van Heyningen, V. (1998). Position effect in human genetic disease. Hum. Mol. Genet. *7*.
4. Milot, E., Fraser, P., and Grosveld, F. (1996). Position effects and genetic disease. Trends Genet. *12*.
5. Fantes, J., Redeker, B., Breen, M., Boyle, S., Brown, J., Fletcher, J., Jones, S., Bickmore, W., Fukushima, Y., Mannens, M., et al. (1995). Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. Hum. Mol. Genet. *4*, 415–422.
6. Bedell, M.A., Brannan, C.I., Evans, E.P., Copeland, N.G., Jenkins, N.A., and Donovan, P.J. (1995). DNA rearrangements located over 100 kb 5′ of the Steel (Sl)-coding region in Steel-panda and Steel-contrasted mice deregulate Sl expression and cause female sterility by disrupting ovarian follicle development. Genes Dev. *9*, 455–70.
7. Wirth, J., Wagner, T., Meyer, J., Pfeiffer, R.A., Tietze, H.-U., Schempp, W., and Scherer, G. (1996). Translocation breakpoints in three patients with campomelic dysplasia and autosomal sex reversal map more than 130 kb from SOX9. Hum. Genet. *97*, 186–193.
8. Krebs, I. (1997). Translocation breakpoint maps 5 kb 3′ from TWIST in a patient affected with Saethre-Chotzen syndrome. Hum. Mol. Genet. *6*, 1079–1086.
9. Flomen, R.H., Vatcheva, R., Gorman, P.A., Baptista, P.R., Groet, J., Barišić, I., Ligutic, I., and Nižetić, D. (1998). Construction and Analysis of a Sequence-Ready Map in 4q25:

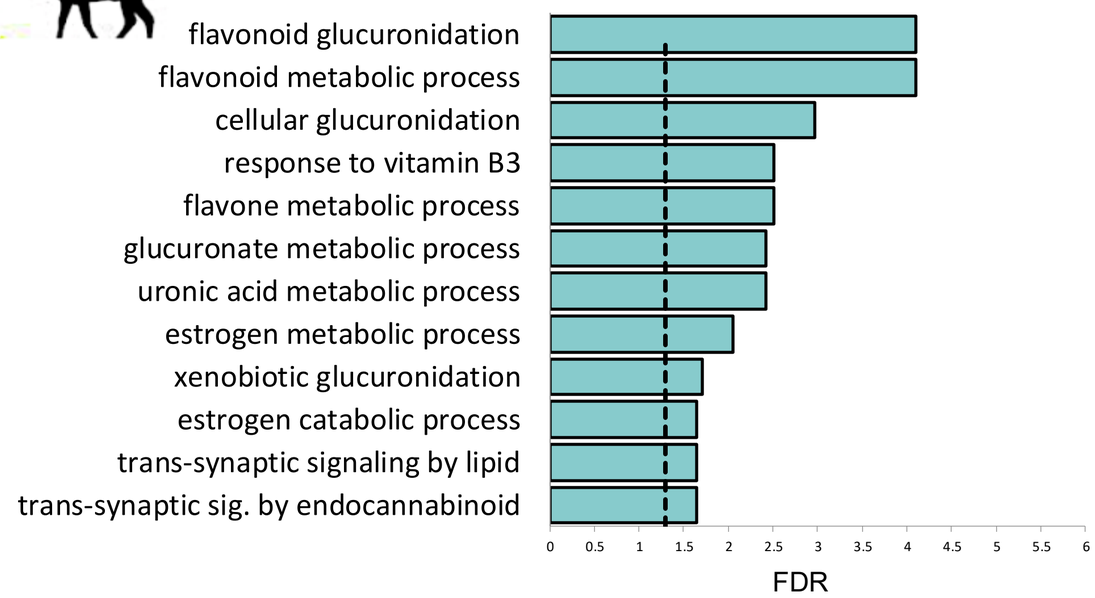
Rieger Syndrome Can Be Caused by Haploinsufficiency ofRIEG, but Also by Chromosome Breaks ≈90 kb Upstream of This Gene. Genomics *47*, 409–413.
10. De, S., Teichmann, S.A., and Babu, M.M. (2009). The impact of genomic neighborhood on the evolution of human and chimpanzee transcriptome. Genome Res. *19*, 785–94.
11. Bagadia, M., Chandradoss, K.R., Jain, Y., Singh, H., Lal, M., and Sandhu, K.S. (2019). Evolutionary Loss of Genomic Proximity to Conserved Noncoding Elements Impacted the Gene Expression Dynamics During Mammalian Brain Development. Genetics *211*, 1239–1254.
12. Real, F.M., Haas, S.A., Franchini, P., Xiong, P., Simakov, O., Kuhl, H., Schöpflin, R., Heller, D., Moeinzadeh, M.H., Heinrich, V., et al. (2020). The mole genome reveals regulatory rearrangements associated with adaptive intersexuality. Science (80-.). *370*.
13. Kim, J., Farré, M., Auvil, L., Capitanu, B., Larkin, D.M., Ma, J., and Lewin, H.A. (2017). Reconstruction and evolutionary history of eutherian chromosomes. Proc. Natl. Acad. Sci. U. S. A. *114*, E5379–E5388.
14. Luo, H., Arndt, W., Zhang, Y., Shi, G., Alekseyev, M.A., Tang, J., Hughes, A.L., and Friedman, R. (2012). Phylogenetic analysis of genome rearrangements among five mammalian orders. Mol. Phylogenet. Evol. *65*, 871–82.
15. Dunn-Fletcher, C.E., Muglia, L.M., Pavlicev, M., Wolf, G., Sun, M.A., Hu, Y.C., Huffman, E., Tumukuntala, S., Thiele, K., Mukherjee, A., et al. (2018). Anthropoid primate–specific retroviral element THE1B controls expression of CRH in placenta and alters gestation length. PLoS Biol. *16*.
16. Bozek, K., Wei, Y., Yan, Z., Liu, X., Xiong, J., Sugimoto, M., Tomita, M., Pääbo, S., Sherwood, C.C., Hof, P.R., et al. (2015). Organization and Evolution of Brain Lipidome Revealed by Large-Scale Analysis of Human, Chimpanzee, Macaque, and Mouse Tissues. Neuron *85*.
17. Spanaki, C., Sidiropoulou, K., Petraki, Z., Diskos, K., Konstantoudaki, X., Volitaki, E., Mylonaki, K., Savvaki, M., and Plaitakis, A. (2024). Glutamate-specific gene linked to human brain evolution enhances synaptic plasticity and cognitive processes. iScience *27*.
18. Muntané, G., Horvath, J.E., Hof, P.R., Ely, J.J., Hopkins, W.D., Raghanti, M.A., Lewandowski, A.H., Wray, G.A., and Sherwood, C.C. (2015). Analysis of synaptic gene expression in the neocortex of primates reveals evolutionary changes in glutamatergic neurotransmission. Cereb. Cortex *25*.
19. Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J.F., Faraut, T., Wu, C., Muzny, D.M., Li, Y., Zhang, W., et al. (2014). The sheep genome illuminates biology of the rumen and lipid metabolism. Science (80-.). *344*.
20. Kawai, Y.K., Sano, K., Ikenaka, Y., Nakayama, S.M.M., Kondo, M., Kubota, A., and Ishizuka, M. (2021). Evolutionary history of mammalian UDP-glucuronosyltransferase (UGT)1 and UGT2 families: the emergence of UGT2B subfamily in eutherians after the diversification of flowering plants. bioRxiv.
21. Wooding, F.B.P., and Burton, G.J. (2008). Comparative placentation: Structures, functions and evolution.
22. Diessler, M., Ventureira, M., Hernandez, R., Sobarzo, C., Casas, L., Barbeito, C., and Cebral, E. (2017). Differential expression and activity of matrix metalloproteinases 2 and 9 in canine early placenta. Reprod. Domest. Anim. *52*.
23. Benjamin, A., Costa, A.N. Da, Delaunois, A., Rosseels, M.L., and Valentin, J.P. (2015).

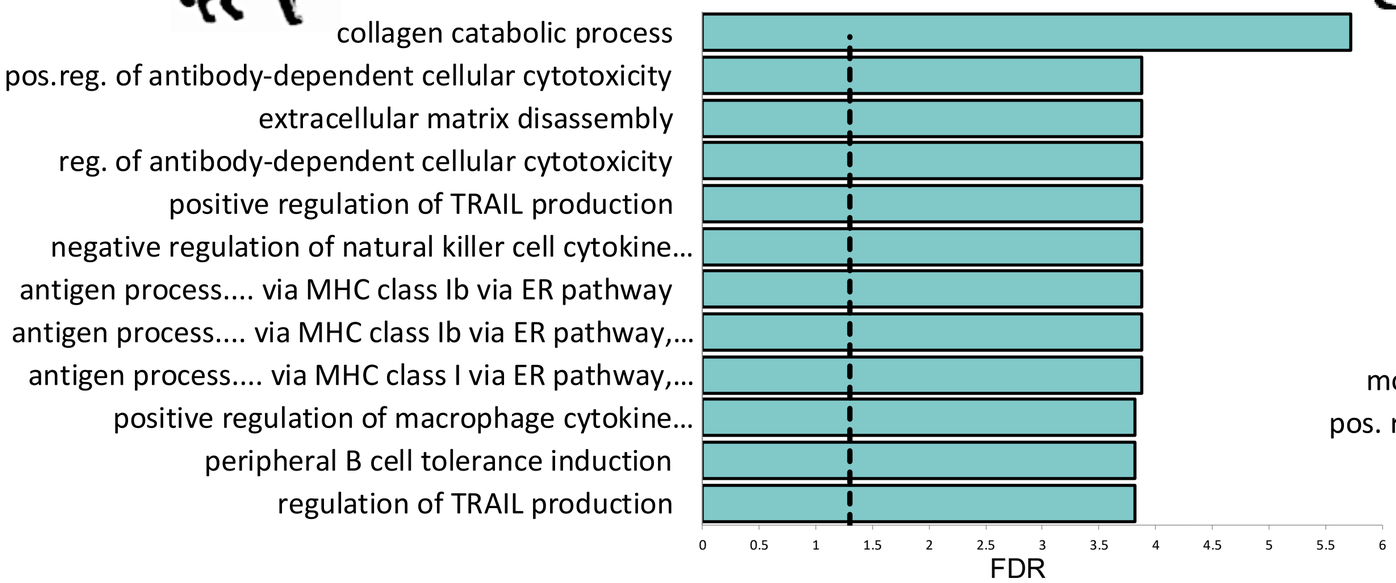
Renal safety pharmacology in drug discovery and development. Handb. Exp. Pharmacol. *229*.
24. Dickinson, H., Moritz, K., Wintour, E.M., Walker, D.W., and Kett, M.M. (2007). A comparative study of renal function in the desert-adapted spiny mouse and the laboratory-adapted C57BL/6 mouse: Response to dietary salt load. Am. J. Physiol. - Ren. Physiol. *293*.
25. Bittner, N.K.J., Mack, K.L., and Nachman, M.W. (2021). Gene expression plasticity and desert adaptation in house mice*. Evolution (N. Y). *75*.
26. Gbadegesin, R.A., Hall, G., Adeyemo, A., Hanke, N., Tossidou, I., Burchette, J., Wu, G., Homstad, A., Sparks, M.A., Gomez, J., et al. (2014). Mutations in the gene that encodes the F-Actin binding protein anillin cause FSGS. J. Am. Soc. Nephrol. *25*.
27. Nishibori, Y., Katayama, K., Parikka, M., Oddsson, Á., Nukui, M., Hultenby, K., Wernerson, A., He, B., Ebarasi, L., Raschperger, E., et al. (2011). Glcci1 deficiency leads to proteinuria. J. Am. Soc. Nephrol. *22*.
28. Tomas, N.M., Beck, L.H., Meyer-Schwesinger, C., Seitz-Polski, B., Ma, H., Zahner, G., Dolla, G., Hoxha, E., Helmchen, U., Dabert-Gay, A.-S., et al. (2014). Thrombospondin Type-1 Domain-Containing 7A in Idiopathic Membranous Nephropathy. N. Engl. J. Med. *371*.
29. Pays, E. (2021). The function of apolipoproteins L (APOLs): relevance for kidney disease, neurotransmission disorders, cancer and viral infection. FEBS J. *288*.
30. Jebb, D., Huang, Z., Pippel, M., Hughes, G.M., Lavrichenko, K., Devanna, P., Winkler, S., Jermiin, L.S., Skirmuntt, E.C., Katzourakis, A., et al. (2020). Six reference-quality genomes reveal evolution of bat adaptations. Nature *583*.
31. Sulak, M., Fong, L., Mika, K., Chigurupati, S., Yon, L., Mongan, N.P., Emes, R.D., and Lynch, V.J. (2016). TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. Elife *5*.
32. Hudson, N.J., Baker, M.L., Hart, N.S., Wynne, J.W., Gu, Q., Huang, Z., Zhang, G., Ingham, A.B., Wang, L., and Reverter, A. (2014). Sensory rewiring in an echolocator: Genome-wide modification of retinogenic and auditory genes in the bat Myotis davidii. G3 Genes, Genomes, Genet. *4*.
33. Wang, H., Zhao, H., Sun, K., Huang, X., Jin, L., and Feng, J. (2019). Evolutionary Basis of High-Frequency Hearing in the Cochleae of Echolocators Revealed by Comparative Genomics. Genome Biol. Evol. *12*.
34. Savage, V.M., Allen, A.P., Brown, J.H., Gillooly, J.F., Herman, A.B., Woodruff, W.H., and West, G.B. (2007). Scaling of number, size, and metabolic rate of cells with body size in mammals. Proc. Natl. Acad. Sci. U. S. A. *104*.
35. Lewis, J.H. (1974). Comparative hematology: Studies on elephants, Elephas maximus. Comp. Biochem. Physiol. -- Part A Physiol. *49*.
36. Guntawang, T., Sittisak, T., Kochagul, V., Srivorakul, S., Photichai, K., Boonsri, K., Janyamethakul, T., Boonprasert, K., Langkaphin, W., Thitaram, C., et al. (2021). Pathogenesis of hemorrhagic disease caused by elephant endotheliotropic herpesvirus (EEHV) in Asian elephants (Elephas maximus). Sci. Rep. *11*.
37. Poole, D.C., and Erickson, H.H. (2011). Highly athletic terrestrial mammals: Horses and dogs. Compr. Physiol. *1*.
38. Doherty, A., and de Magalhães, J.P. (2016). Has gene duplication impacted the evolution of Eutherian longevity? Aging Cell *15*.
39. Derous, D., Sahu, J., Douglas, A., Lusseau, D., and Wenzel, M. (2021). Comparative

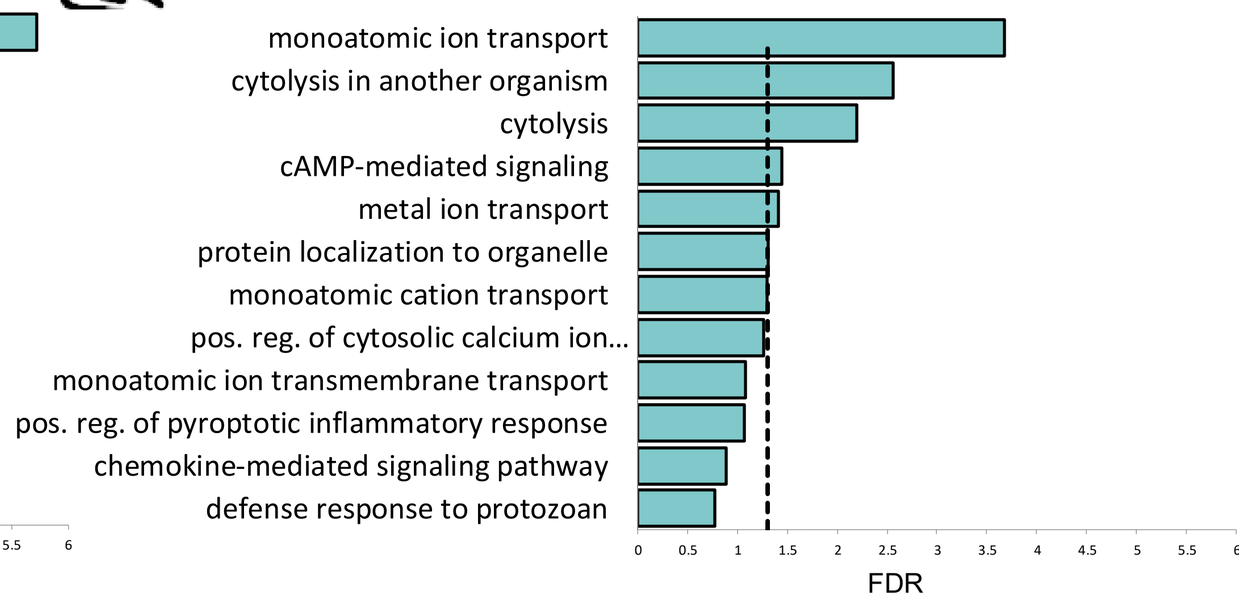
genomics of cetartiodactyla: Energy metabolism underpins the transition to an aquatic lifestyle. Conserv. Physiol. *9*.
40. Jezova, D., Ochedalski, T., Glickman, M., Kiss, A., and Aguilera, G. (1999). Central corticotropin-releasing hormone receptors modulate hypothalamic- pituitary-adrenocortical and sympathoadrenal activity during stress. Neuroscience *94*.
41. Smith, R., and Nicholson, R.C. (2007). Corticotrophin releasing hormone and the timing of birth. Front. Biosci. *12*.
42. Pusey, A., Williams, J., and Goodall, J. (1997). The influence of dominance rank on the reproductive success of female chimpanzees. Science (80-.). *277*.
43. Sapolsky, R.M. (2021). Glucocorticoids, the evolution of the stress-response, and the primate predicament. Neurobiol. Stress *14*.
44. Akhtar, W., de Jong, J., Pindyurin, A. V, Pagie, L., Meuleman, W., de Ridder, J., Berns, A., Wessels, L.F.A., van Lohuizen, M., and van Steensel, B. (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell *154*, 914–27.
45. Chen, X., and Zhang, J. (2016). The Genomic Landscape of Position Effects on Protein Expression Level and Noise in Yeast. Cell Syst. *2*, 347–54.
46. Singh, A., Bagadia, M., and Sandhu, K.S. (2016). Spatially coordinated replication and minimization of expression noise constrain three-dimensional organization of yeast genome. DNA Res *23*, 155–169.
47. Chiang, C., Scott, A.J., Davis, J.R., Tsang, E.K., Li, X., Kim, Y., Hadzic, T., Damani, F.N., Ganel, L., GTEx Consortium, et al. (2017). The impact of structural variation on human gene expression. Nat. Genet. *49*, 692–699.
48. Xu, Z., Lee, D.-S., Chandran, S., Le, V.T., Bump, R., Yasis, J., Dallarda, S., Marcotte, S., Clock, B., Haghani, N., et al. (2022). Structural variants drive context-dependent oncogene activation in cancer. Nature *612*, 564–572.
49. Liu, W., Zhu, P., Li, M., Li, Z., Yu, Y., Liu, G., Du, J., Wang, X., Yang, J., Tian, R., et al. (2023). Large-scale across species transcriptomic analysis identifies genetic selection signatures associated with longevity in mammals. EMBO J. *42*.
50. Laverré, A., Tannier, E., and Necsulea, A. (2022). Long-range promoter–enhancer contacts are conserved during evolution and contribute to gene expression robustness. Genome Res. *32*.

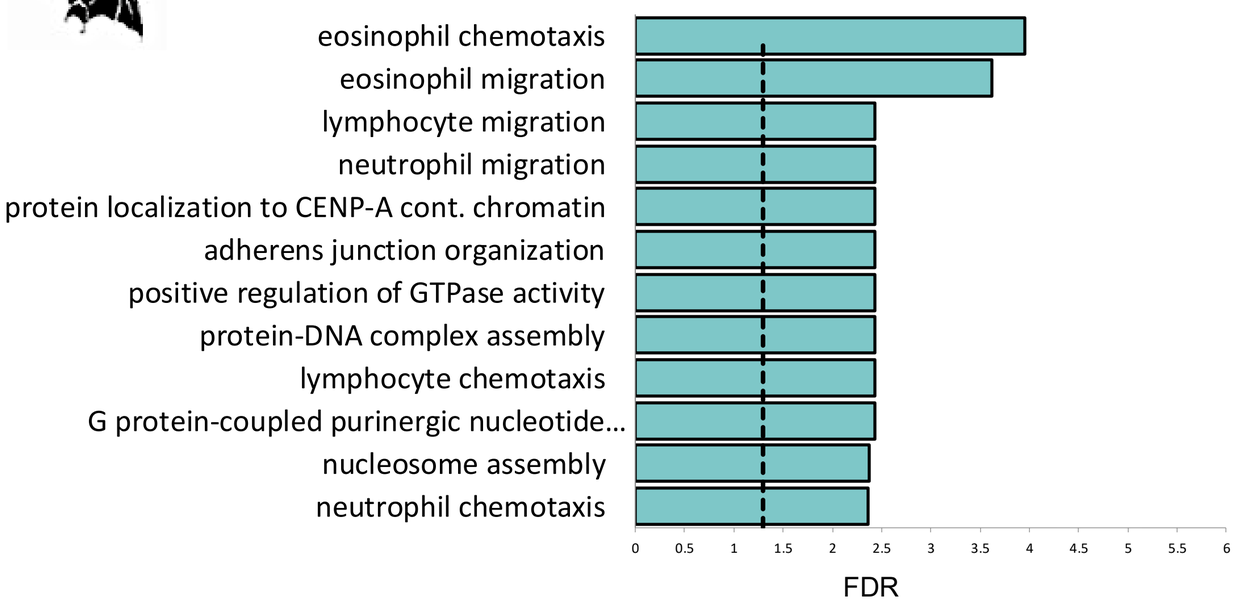

a
2mb
5mb
10Mb
2 3 4 5 6 7 8
Inter-genic distance (log10)
Species 2
cis split
>10Mb
trans split
Species 2
Species 1
<1Mb
<1Mb
Clade-specific LoS
b
Platypus**
Primates
Rodentia
Cetacean
Artiodactyla
Perissodactyla
Carnivora
Chiroptera
Afrotheria
Tasmanian devil**
<= 1mb
8 High GR
6
4
2
0 Low GR
(log2)
**Platypus
Primates
Rodentia
Cetacean
Artiodactyla
Perissodactyla
Carnivora
Chiroptera
Afrotheria
Tasmanian devil**
> 10mb / trans
c
n= 12036
LOSS
Without breakpoint
With breakpoint

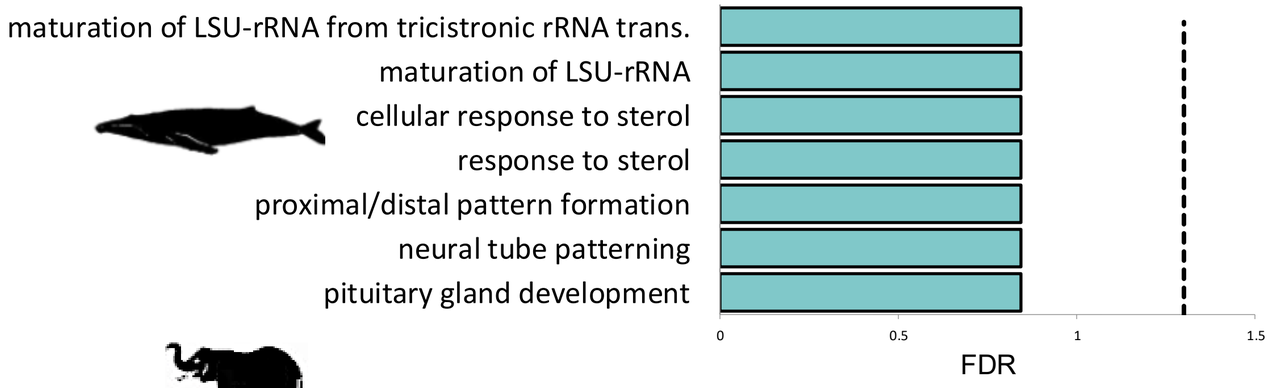

glucocorticoid secretion
corticosterone secretion
est. of planar polarity inv.in nephron morpho.
regulation of glucocorticoid secretion
corticosteroid hormone secretion
regulation of corticosterone secretion
lipid export from cell
neg. reg. of synaptic transmission
sodium-independent organic anion transport
positive regulation of secretion by cell
reg. of corticosteroid hormone secretion
behavior
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6
FDR

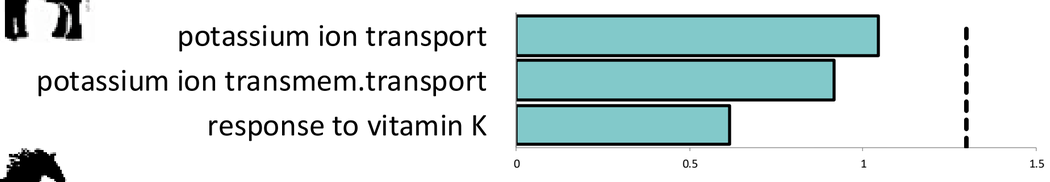

flavonoid glucuronidation
flavonoid metabolic process
cellular glucuronidation
response to vitamin B3
flavone metabolic process
glucuronate metabolic process
uronic acid metabolic process
estrogen metabolic process
xenobiotic glucuronidation
estrogen catabolic process
trans-synaptic signaling by lipid
trans-synaptic sig. by endocannabinoid
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6
FDR

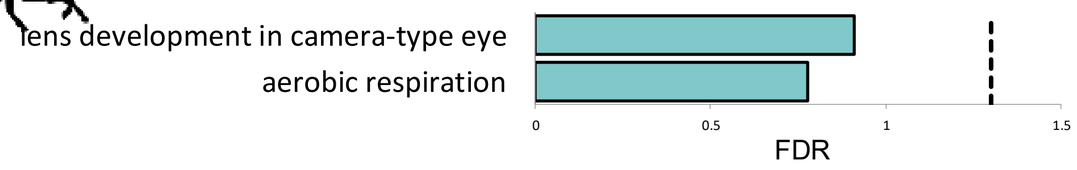

collagen catabolic process
pos.reg. of antibody-dependent cellular cytotoxicity
extracellular matrix disassembly
reg. of antibody-dependent cellular cytotoxicity
positive regulation of TRAIL production
negative regulation of natural killer cell cytokine...
antigen process.... via MHC class Ib via ER pathway
antigen process.... via MHC class Ib via ER pathway,...
antigen process.... via MHC class I via ER pathway,...
positive regulation of macrophage cytokine...
peripheral B cell tolerance induction
regulation of TRAIL production
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6
FDR

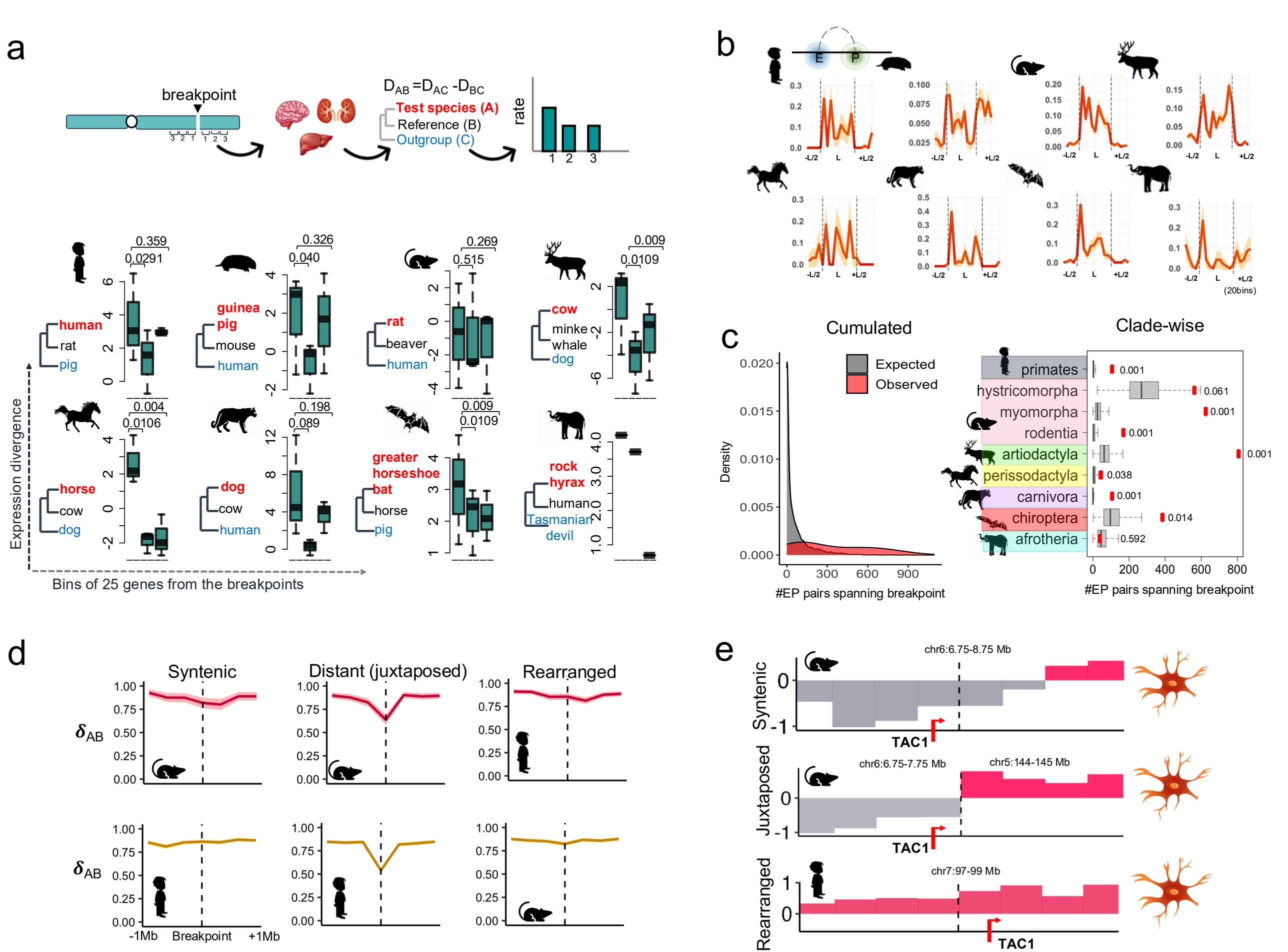

monoatomic ion transport
cytolysis in another organism
cytolysis
cAMP-mediated signaling
metal ion transport
protein localization to organelle
monoatomic cation transport
pos. reg. of cytosolic calcium ion...
monoatomic ion transmembrane transport
pos. reg. of pyroptotic inflammatory response
chemokine-mediated signaling pathway
defense response to protozoan
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6
FDR

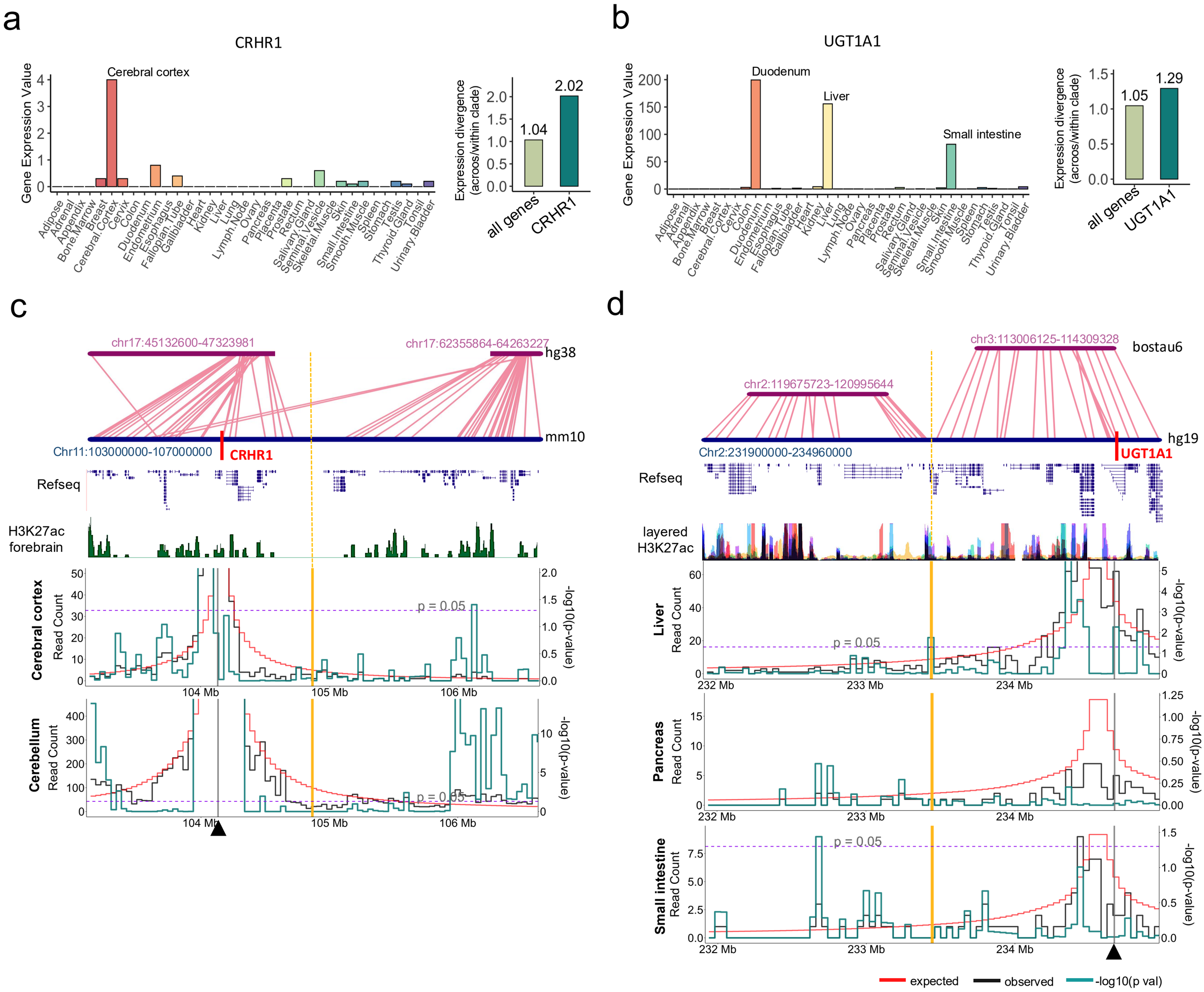

eosinophil chemotaxis
eosinophil migration
lymphocyte migration
neutrophil migration
protein localization to CENP-A cont. chromatin
adherens junction organization
positive regulation of GTPase activity
protein-DNA complex assembly
lymphocyte chemotaxis
G protein-coupled purinergic nucleotide...
nucleosome assembly
neutrophil chemotaxis
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6
FDR

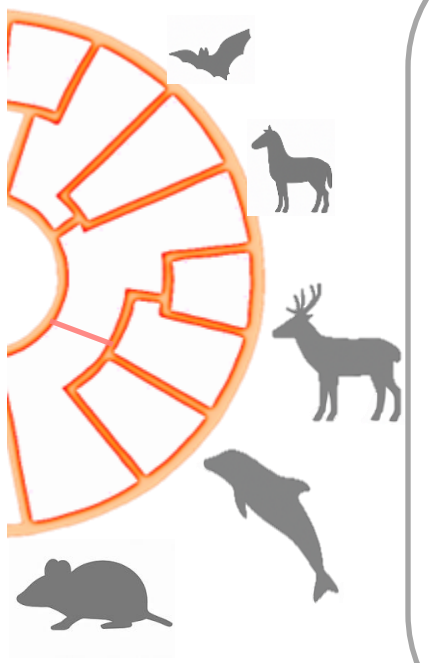

maturation of LSU-rRNA from tricistronic rRNA trans.
maturation of LSU-rRNA
cellular response to sterol
response to sterol
proximal/distal pattern formation
neural tube patterning
pituitary gland development
0 0.5 1 1.5
FDR


potassium ion transport
potassium ion transmem.transport
response to vitamin K
0 0.5 1 1.5

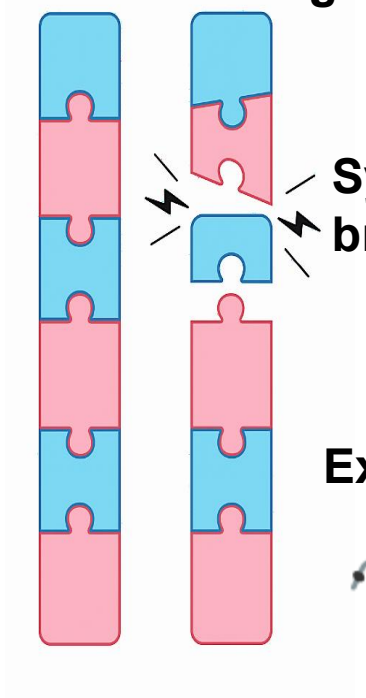

lens development in camera-type eye
aerobic respiration
0 0.5 1 1.5
FDR

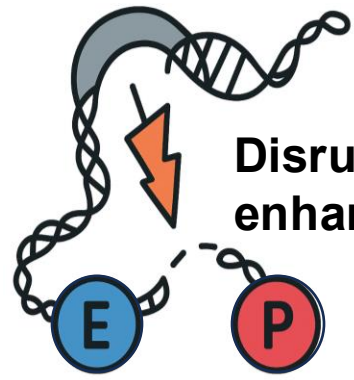

a
breakpoint
$D_{AB} = D_{AC} - D_{BC}$
Test species (A)
Reference (B)
Outgroup (C)
rate
Expression divergence
Bins of 25 genes from the breakpoints
human
rat
pig
guinea pig
mouse
human
rat
beaver
human
cow
minke whale
dog
horse
cow
dog
dog
cow
human
greater horseshoe bat
horse
pig
rock hyrax
human
Tasmanian devil
b
E
P
(20bins)
c
Cumulated
Clade-wise
Expected
Observed
Density
#EP pairs spanning breakpoint
primates
hystricomorpha
myomorpha
rodentia
artiodactyla
perissodactyla
carnivora
chiroptera
afrotheria
d
Syntenic
Distant (juxtaposed)
Rearranged
δ_{AB}
-1Mb
Breakpoint
+1Mb
e
chr6:6.75-8.75 Mb
chr6:6.75-7.75 Mb
chr5:144-145 Mb
chr7:97-99 Mb
TAC1


a
CRHR1
Cerebral cortex
Gene Expression Value
Expression divergence (acroos/within clade)
1.04
2.02
all genes
CRHR1
b
UGT1A1
Duodenum
Liver
Small intestine
1.05
1.29
all genes
UGT1A1
c
chr17:45132600-47323981
chr17:62355864-64263227
hg38
mm10
Chr11:103000000-107000000
CRHR1
Refseq
H3K27ac forebrain
Cerebral cortex
Cerebellum
Read Count
-log10(p-value)
p = 0.05
104 Mb
105 Mb
106 Mb
d
chr3:113006125-114309328
bostau6
chr2:119675723-120995644
hg19
Chr2:231900000-234960000
UGT1A1
Refseq
layered H3K27ac
Liver
Pancreas
Small intestine
232 Mb
233 Mb
234 Mb
expected
observed
-log10(p val)



Genomic rearrangements

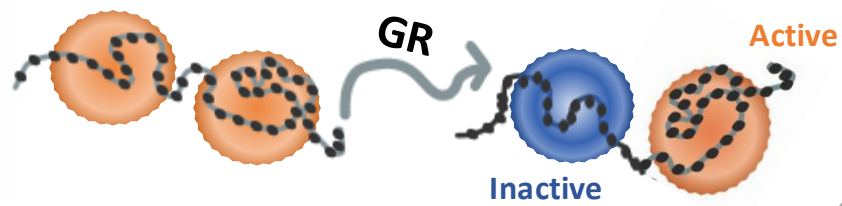


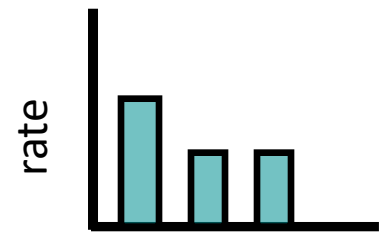


Disruption of enhancer contacts

Position effects

Exchange of A/B chromatin domains

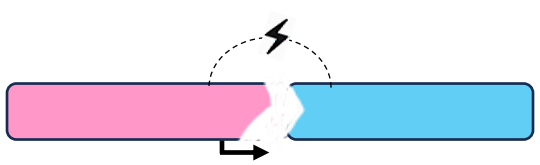




Expression divergence

Downstream effects

Example-



Primates-specific position effect

↓

Adaptation to stress ?