Published in final edited form as:
Cell Rep. 2018 March 06; 22(10): 2742–2755. doi:10.1016/j.celrep.2018.02.008.

Accelerated Evolution in Distinctive Species Reveals Candidate Elements for Clinically Relevant Traits, Including Mutation and Cancer Resistance

Elliott Ferris[1], **Lisa M. Abegglen**[2,5], **Joshua D. Schiffman**[2,3,5], and **Christopher Gregg**[1,4,6,7,*]

[1]Department of Neurobiology and Anatomy, University of Utah, Salt Lake City, UT 84132-3401, USA

[2]Department of Pediatrics, University of Utah, Salt Lake City, UT 84132-3401, USA

[3]Department of Oncological Sciences, University of Utah, Salt Lake City, UT 84132-3401, USA

[4]Department of Human Genetics, University of Utah, Salt Lake City, UT 84132-3401, USA

[5]Huntsman Cancer Institute, Salt Lake City, UT, USA

[6]New York Stem Cell Foundation, New York, NY, USA

SUMMARY

The identity of most functional elements in the mammalian genome and the phenotypes they impact are unclear. Here, we perform a genomewide comparative analysis of patterns of accelerated evolution in species with highly distinctive traits to discover candidate functional elements for clinically important phenotypes. We identify accelerated regions (ARs) in the elephant, hibernating bat, orca, dolphin, naked mole rat, and thirteen-lined ground squirrel lineages in mammalian conserved regions, uncovering ~33,000 elements that bind hundreds of different regulatory proteins in humans and mice. ARs in the elephant, the largest land mammal, are uniquely enriched near elephant DNA damage response genes. The genomic hotspot for elephant ARs is the E3 ligase subunit of the Fanconi anemia complex, a master regulator of DNA repair. Additionally, ARs in the six species are associated with specific human clinical phenotypes that have apparent concordance with overt traits in each species.

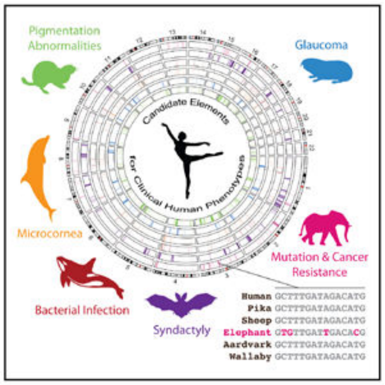
Graphical Abstract

[*]Correspondence: chris.gregg@neuro.utah.edu.
[7]Lead Contact

AUTHOR CONTRIBUTIONS
Conceptualization, E.F. and C.G.; Methodology, E.F. and C.G.; Investigation, E.F., L.M.A, J.S., and C.G.; elephant cell DNA damage RNA-seq experiment, L.M.A. and J.D.S.; Software, E.F. and C.G.; Formal Analysis, E.F. and C.G.; Data Visualization: E.F. and C.G.; Writing, C.G. and E.F.; Supervision, C.G. and J.S.; Funding, C.G. and J.D.S.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures, five figures, and seven tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.02.008.

DECLARATION OF INTERESTS
The authors declare no competing interests.



INTRODUCTION

Over 5,400 different mammalian species exist globally and some evolved highly distinctive phenotypic traits. The African elephant, for instance, is the largest land mammal, with a body mass up to 7,000 kg. The bat is the only aerial mammalian lineage, and other species have unique adaptations for marine environments (e.g., orca and dolphin) or subterranean environments (e.g., naked mole rat). Seasonal scarcities in resource availability have also driven the evolution of hibernation in some species. For example, the little brown bat, big brown bat, and thirteen-lined ground squirrel are obligate hibernators (Carey et al., 2003). The genomic mechanisms underlying these phenotypes are largely unknown but could be relevant for understanding human disease. For example, cancer risk is strongly influenced by cell division frequency and DNA replication errors (Vassilev and DePamphilis, 2017; Zhu et al., 2016), and Peto's paradox predicts that large, long-lived mammals, such as elephants, evolved unique mechanisms to reduce the risk for cancer-causing somatic mutations (Caulin and Maley, 2011; Tollis et al., 2017). We and others recently identified an expanded number of *TP53* gene copies in the elephant genome, indicating one candidate mechanism for increased cancer resistance (Abegglen et al., 2015; Sulak et al., 2016). Currently however, we do not fully understand the mechanisms for improved cancer prevention in the elephant or the mechanisms underlying distinctive phenotypes in most mammalian species.

New phenotypes frequently arise due to evolutionary changes to noncoding regulatory elements rather than protein-coding changes (Carroll, 2008; Wray, 2007). Although much of the genome is biochemically active (ENCODE Project Consortium, 2012), identifying functional elements for particular traits is challenging, and the best approaches are debated (Kellis et al., 2014). One approach is to focus on conserved genomic regions. Indeed, species-specific changes to conserved noncoding elements are linked to some major phenotypic effects, such as the loss of limbs in the snake (Kaltcheva and Lewandoski, 2016; Kvon et al., 2016) and the loss of penile spines in humans (McLean et al., 2011). Conserved elements exhibiting accelerated evolution in a particular species may have roles in shaping the traits of that species (Bird et al., 2007; Boyd et al., 2015; Capra et al., 2013; Hubisz et al., 2011; Kim and Pritchard, 2007; Lindblad-Toh et al., 2011; Pollard et al., 2006a, 2006b, 2010; Prabhakar et al., 2006). Accelerated regions (ARs) are best known from studies of human ARs and are conserved elements with significantly increased nucleotide substitution rates due to the effects of positive selection, relaxed purifying selection, or GC-biased gene conversion in a particular lineage (Hubisz and Pollard, 2014; Kostka et al., 2012; Pollard et al., 2010). For example, one human AR is an enhancer with putative roles in the evolution of the human thumb (Prabhakar et al., 2008). Despite these advances, the identity and roles of most functional elements in the mammalian genome remain unclear.

Here, we reasoned that a comparative genomic analysis of ARs in species with distinctive traits could facilitate the discovery of conserved functional elements that changed to shape those unique traits. We performed a comparative, genome-wide analysis of accelerated evolution in the African elephant, little brown bat (microbat), big brown bat, orca, bottlenose dolphin (dolphin), naked mole rate (mole rat), and thirteen-lined ground squirrel (squirrel) (Figures 1A and 1B). These species represent mammalian adaptations to different environments (terrestrial, aerial, aquatic, and subterranean). They also have distinctive physiological and anatomical traits, including massive body size (elephant), wings (bat), and fins and flukes (dolphin and orca), hibernation (squirrel, microbat, and big brown bat), and adaptations for diving (orca and dolphin). Our study identifies ARs in each target species in genomic elements conserved across other mammals. We uncover high-priority, candidate functional elements in the mammalian genome for traits of biomedical importance, such as mutation and cancer resistance in elephants.

RESULTS

Comparative Genome-wide Analysis of Accelerated Evolution Reveals Candidate Functional Genomic Elements for Distinctive Mammalian Traits

We began with a comparative, genomewide analysis of accelerated evolution in the elephant, microbat, big brown bat, orca, dolphin, mole rat, and squirrel. To facilitate comparisons between species, we created a common background of conserved genomic regions using available genomes of 20 mammalian species ranging from marsupials to humans (Figure 1B; Supplemental Experimental Procedures), and identified 50-bp genomic elements present in our target species and conserved across our background species using PhastCons (Hubisz et al., 2011). The analysis uncovered 660,851 mammalian conserved autosomal regions. This amounts to ~1% of the human genome. We used RPhast (Hubisz et al., 2011) to test for these regions for statistically significant increases in nucleotide substitution rates compared to the neutral model in each of the target species (5% false discovery rate [FDR]). Our analysis identified 3,458 elephant ARs (0.5% of conserved regions), 18,000 microbat ARs (2.7%), 19,152 big brown bat ARs (2.9%), 2,608 orca ARs (0.4%), 2,408 dolphin ARs (0.4%), 4440 mole rat ARs (0.7%), and 6,704 squirrel ARs (1.0%) (Figure 1C; Table S1). The ARs are located across all human autosomes (Figures 1 and S1A). Most ARs are species specific. There are small, but significant AR overlaps between some species (Figure 1C), indicating some common evolutionary patterns.

Sequence quality could impact AR discovery. We performed control experiments to test for reproducibility, particularly for the bats, which have large numbers of ARs. A comparison of microbat ARs (*Myotis lucifugus*) to ARs in the closely related *Myotis davidii* species found that 14,331 (68%) ARs are shared (Figures 1 and S1B). A comparison of the microbat to a more distantly related hibernating bat, the big brown bat, uncovered 13,665 shared ARs

(58%) (Figures 1C and S1B). Therefore, the majority of bat ARs are reproducible between close lineages. The shared microbat-big brown bat ARs are candidate elements for shaping phenotypes in the hibernating bat lineage. We focus on these “hibernating bat” (Hib bat) ARs in the remainder of our study (Figure 1C). A large fraction of ARs (37%) are also reproducible between the closely related dolphin and orca (Figure 1C). Therefore, more ARs are shared between more closely related species, which is expected only if the data are not dominated by the effects of poor genome sequences. A significant linear relationship (r = 0.9, p = 0.0041, linear regression) exists between the number of ARs in a species and the phylogenetic distance between that species and their closest relative in the background (Figures 1 and S1C). Significant relationships between AR number and assembly quality (N50) or assembly size were not observed (data not shown). Thus, our ARs are promising candidates for shaping species-specific traits.

We found that 18%–35% of the species’ ARs are located in human genome annotated exons, while the majority (65%–82%) are noncoding regions (Figure 1D). The ARs are in otherwise conserved elements predicted to be functional across mammalian species. We tested whether the human and mouse homologs of the ARs show evidence of functionality by comparing AR regions to data for 6,387 human and 5,827 mouse chromatin immunoprecipitation (ChIP) datasets in the ChIP-Atlas database (Table S2). The results show significant enrichments for hundreds of transcription factor (TF) and regulatory-protein-binding sites (Figures 2A and 2B; FDR < 5% relative to random genomic elements). A comparison of the ARs to DNAse I hypersensitivity site sequencing (DNase-seq) datasets for 19 human and 14 mouse cell types (Table S2) uncovered significant enrichments for DNase-I-hypersensitive regions in various human (Figure 2C) and mouse (Figure 2D) cell types (FDR < 5%).

It is estimated that ~30% of human ARs are enhancers (Capra et al., 2013) and some are noncoding RNAs, but most remain uncharacterized (Hubisz and Pollard, 2014). We investigated the nature of the human and mouse homologs of the species’ ARs by comparing these regions to available ChIP sequencing (ChIP-seq) and genomics datasets for 68 and 76 epigenetic marks in human and mouse cells, respectively, including various histone modifications, 5-mC methylated DNA, 5-hmC hydroxymethylated DNA, CTCF (CCCTC-binding factor)-binding sites and EP300 (E1A-binding protein p300)-binding sites (Table S2, ChIP-Atlas). Human (Figure 3A) and mouse (Figure 3B) homologs of the species’ ARs are significantly enriched for 34 and 44 different epigenetic marks, respectively (FDR < 5% relative to random elements). Significant enrichments for markers of active enhancers (H3K27ac, H3K4me2, H3K4me1, and EP300), active promoters (H3K4me3 and H3K4me2), transcribed elements (H3K36me3), and repressed elements (H3K27me3) were observed for the human and mouse homologs of all six species’ ARs (Figures 3A and 3B; Table S2), suggesting functional elements impacted by ARs across different species. We do not know whether these results extend beyond the mouse and human homologs.

We also found apparent species differences. CTCF boundary elements are enriched at human (Figure 3A) and mouse (Figure 3B) homologs for elephant, dolphin, and mole rat ARs, but not for orca, Hib bat, or squirrel ARs. In total, 26 and 31 classes of biochemically active human and mouse elements, respectively, show significant enrichments for ARs from some species, but not others (Figures 3A and B). Therefore, ARs in some lineages impact relatively more elements of a particular class than ARs in other lineages. This may help reveal how particular phenotypes evolve. Note that the human and mouse ChIP experiments are not matched by cell type or time point, and therefore, we cannot compare the human and mouse results. Overall, we found a composite of candidate functional elements for shaping biological processes in various cell types. We refine these findings below to uncover candidate regions for specific phenotypes.

A Genomic Hotspot for Elephant ARs Uncovers a Putative Gene Module for Mutation and Cancer Resistance

We hypothesize that the distinctive target species phenotypes can help decipher the function of different ARs. The elephant is postulated to have unique mechanisms for enhanced mutation and cancer resistance to manage the large number of cell divisions required to grow and maintain its body size (Abegglen et al., 2015; Caulin and Maley, 2011). Therefore, in a proof-of-principle analysis, we tested whether elephant ARs uniquely reveal candidate elements for shaping mutation and cancer resistance phenotypes compared to other species' ARs.

ARs were assigned to genes according to hg19 annotations using a custom algorithm based on GREAT (default basal plus extension settings) (McLean et al., 2010). We determined that the median number of ARs per gene, for genes with ARs, is only ~1–3. However, a few genes are associated with many ARs, indicating hotspots (Figure 4A). Remarkably, the gene associated with most elephant ARs is *FANCL* (289 ARs; Figure 4A; Table S6), a RING domain protein mediating the E3 ligase activity of the Fanconi anemia (FA) core complex. The FA pathway is a master regulator of DNA repair that guards the genome against mutations through homologous recombination, nucleotide excision repair and mutagenic translesion synthesis (Moldovan and D'Andrea, 2009). The enrichment for elephant ARs at this locus is statistically significant and species specific (Figure 4B; $p < 0.0001$, two-tailed chi-square test). *ZNF521* and *NFIA* are hotspots for Hib bat and mole rat ARs, respectively, while *TENM3* and *FOXP1* are hotspots for multiple species' ARs (Figure 4B). The *FANCL* locus is associated with the largest number of conserved regions out of all genes with ARs (Figure 4C; Table S6), revealing high evolutionary constraint at this locus. The lineage specificity of our results suggests that AR hotspots are not simply a function of the number of conserved regions at a locus.

A BLAST-like alignment tool (BLAT) search of the human *FANCL* mRNA sequence against the elephant genome uncovered one hit; *FANCL* gene duplications and pseudogenes are not apparent in the loxAfr3 genome build. Further, the synteny of the genomic region harboring *FANCL* in the human genome appears preserved in the elephant (Figure 4D). In a control study to test whether the enrichment for elephant ARs at the *FANCL* locus is robust to different genome alignments, we identified elephant ARs from the available mm10-rooted multiple alignment file, which has some differences in background species and includes the rock hyrax, a species more closely related to the elephant (Figures 4 and S1A) than the manatee in the hg19-rooted alignment (Figure 1A). Although the alignment and different species influenced the number and identity of the elephant ARs, we nonetheless identified

1,140 elephant ARs common to both the hg19 and mm10 rooted alignments (Figures 4 and S1B). The *FANCL* locus is associated with the largest number of these shared ARs (Figures 4 and S1C). Thus, this hotspot is robust to different alignments and background species.

The genes with the second and seventh largest numbers of elephant ARs are *BCL11A* (B cell lymphoma/leukemia 11A; 281 ARs) and *VRK2* (vaccine-related kinase 2; 55 ARs), respectively, which reside next to *FANCL* in the human and elephant genomes (Figure 4E). The elephant ARs assigned to *VRK2, FANCL* and *BCL11a* are shown in Figure 4E, and reveal that elephant ARs encompasses all three genes and two lncRNAs. VRK2 (Blanco et al., 2006; Klerkx et al., 2009) and BCL11A (Yu et al., 2012) are regulators of p53, which has an expanded copy number in the elephant genome (Abegglen et al., 2015; Sulak et al., 2016). Vaccinia-related kinases (VRK1 and VRK2) phosphorylate p53 on Thr-18, promoting p53 stability and activity and inhibiting interactions with MDM2, which regulates p53 degradation (Blanco et al., 2006; Klerkx et al., 2009). In contrast, BCL11A is a proto-oncogene with the opposite effect, negatively regulating p53, promoting increased MDM2 expression, cell proliferation, and survival (Yu et al., 2012; Khaled et al., 2015). Thus, the *VRK2-FANCL-BCL11A* gene block is an important candidate locus for shaping mutation and cancer resistance (Figure 4E, inset). The functions of the lncRNAs (*LINC01122* and *LINC01793*) and microRNAs in this region are unknown. We tested whether the elephant ARs in this hotspot are best explained by adaptive effects (positive selection or relaxed purifying selection) or GC-biased gene conversion (Kostka et al., 2012). We found that the nucleotide substitution patterns are best explained by adaptive effects, because the adaptive model has a greater log likelihood (log-likelihood adaptive model – log-likelihood GC-biased conversion model = +665.8). We further compared a model of adaptive effects + GC-biased conversion to a nested model of GC-biased conversion alone. The model including adaptive effects has a significantly greater log-likelihood ($p < 1 \times 10^{-100}$; chi-square likelihood ratio test; log likelihood difference = +870; Experimental Procedures). Thus, a top hotspot for elephant ARs involves adaptive effects at a master regulator of DNA repair, consistent with predictions from Peto’s paradox (Caulin and Maley, 2011).

Accelerated Evolution in the Elephant Changed Conserved Regulatory Motifs Associated with the *VRK2-FANCL-BCL11A* Genomic Locus

All but three of the elephant ARs at the *VRK2-FANCL-BCL11A* genomic locus impact non-coding elements (Figure 4E). By comparing these to DNase-seq and ChIP-seq datasets for different human cell types in the ChIP-Atlas, we found that the human homologs are significantly enriched for elements active in developing neural cells (Figures 5A and 5B). We found a 10-fold enrichment for elephant AR homologs in DNase-I-hypersensitive sites in human fetal brain (FDR = 4×10^{-5} relative to random elements), an 8-fold enrichment for H3K27ac sites in embryonic stem cell-derived neural precursor cells (NPCs) (FDR = 0.008), and a 23-fold enrichment for SOX2-binding sites in NPCs (FDR = 0.008) (Figures 5A–5C; Table S2). Significant enrichments in other cell or tissue types were not observed. Intriguingly, SOX2 is a critical regulator of stem cell pluripotency, self-renewal, and differentiation (Abdelalim et al., 2014). The *SOX2* locus itself is the 8th most enriched genomic site for elephant ARs (Table S6). These findings suggest SOX2- mediated gene regulation and *VRK2-FANCL-BCL11A* elements active in neural cell lineages changed in the elephant.

To test whether the coding and noncoding genes in the *VRK2-FANCL-BCL11A* locus are expressed in brain and exhibit evidence for shared regulatory effects, we examined the developmental expression patterns of these genes in the Human BrainSpan Atlas (Miller et al., 2014) and compared them to *SOX2* and *TP53* (Figures 5 and S1A). These genes show increased expression during early stages of brain development (Figures 5 and S1A). Across the 524 brain regions and ages in the atlas, *FANCL* expression is significantly correlated to the expression of its neighboring genes, as well as to *SOX2* and *TP53* (Figures 5 and S1B; p < 0.0001, Pearson's correlation), suggesting some shared regulation. In further support of the *VRK2-FANCL-BCL11A* locus as a block of partially co-regulated genes, it is contained within a single topologically associated domain in the human genome (Figure 5C) (Harmston et al., 2017).

To learn how *VRK2-FANCL-BCL11A* elephant ARs changed regulatory architecture (Figure 4C), we identified previously characterized transcription-factor-binding site (TFBS) motifs (Jolma et al., 2013) in the elephant genome sequence and compared them to orthologous sequences from manatee (closest relative of the elephant in the hg19-rooted alignment), dog (a Laurasiatherian species), and human (a Euarchontoglires species) (Figure 1B). We tallied the number of TFBS motifs present in the manatee, dog, and human sequences (conserved motifs) and the number in the elephant sequence to compare the prevalence of 709 different motifs. Some TFBS motifs are more prevalent than others. While the patterns of sequence evolution in different lineages are not well defined, we at least partially account for motif abundance by normalizing the number of conserved sites lost or new sites gained in the elephant for a given motif to the mean number of sites in the other 3 species (Table S3). This analysis yields normalized "lost-motif" and "gained-motif" scores (Figure 5D). Examples of manatee-dog-human conserved TFBS motifs with high lost-motif scores in the elephant included NR4A2, YY1, and NFIA. In general, many conserved TFBS motifs were lost in the elephant ARs (Figure 5D; Table S3). The gainedmotif score analysis revealed TFBS motifs that likely arose in the elephant ARs, including GSX2, GSX1, and HOXB3 (Figure 5D; Table S3). Overall, elephant ARs at the *VRK2-FANCL-BCL11A* locus changed conserved TFBS motifs, potentially changing regulatory architecture compared to smaller species.

Elephant ARs Are Enriched at Genes that Respond to DNA Damage in Primary Elephant Lymphocytes

To further test whether distinctive species' traits help reveal candidate elements for those traits, and further identify candidates for somatic mutation and cancer prevention, we determined whether elephant ARs are uniquely associated with DNA damage response mechanisms. Primary peripheral blood lymphocytes were isolated from an adult female African elephant and treated with gamma radiation (Gy2.0), damaging the DNA (Figure 6A). RNA was collected at 1, 5, and 24 hr after acute radiation treatment, as well as from untreated control cells. RNA sequencing (RNA-seq) analysis revealed that the number of genes differentially expressed between irradiated and control elephant cells increased from 2 to 435 to 2,338 genes at the 1-, 5-, and 24-hr time points, respectively, after radiation treatment (5% FDR; Figure 6B; Table S4). Gene ontology analysis found that the upregulated genes at the 5-hr time point are significantly enriched for the p53 signaling pathway. Downregulated genes are involved in mitosis, including genes with cytoskeletal, membrane, and phosphoprotein functions (Figure 6C). Genes upregulated at the 24-hr time point are enriched for inflammatory response, alternative splicing, innate immunity, and angiogenesis and involve genes that function as phosphoproteins and mediate vascular endothelial growth factor (VEGF) signaling, transcription, and other signaling processes (Figure 6D). The genes downregulated at the 24-hr point are enriched for acetylation functions, alternative splicing, ribosomal and metabolic functions, viral transcription, nonsensemediated decay, and DNA repair (Figure 6D). Thus, we identified an intrinsic DNA damage gene expression response program in elephant cells.

The 5-hr (Figure 6E) and 24-hr (Figure 6F) DNA damage response genes are significantly enriched for elephant ARs relative to the background of conserved regions. At the 5-hr time point, the odds ratio for a DNA damage response gene to be linked to an elephant AR is at least 10-fold greater than those from the other species (Figure 6E). At the 24-hr time point, the odds ratio is also greatest for elephant ARs (Figure 6F). Notably, the orca is the only other species with significant AR enrichments at the DNA damage genes (Figures 6E and 6F). The orca is also the only other species in our study with a very large body size and is therefore predicted to have improved mutation and cancer resistance (Caulin and Maley, 2011). This may explain the unique elephant and orca AR enrichments at DNA damage response genes.

We tested whether elephant ARs identified in both the hg19- rooted and mm10-rooted genome alignments are also enriched near DNA damage response genes (Figures 4 and S1B). Indeed, these ARs are even more strongly enriched at both the 5 hr (odds ratio = 40; $p < 1 \times 10^{-100}$; LOLA enrichment) and 24 hr (odds ratio = 1.6; $p < 1 \times 10^{-5}$) genes than the hg19-rooted ARs described above (~2-fold stronger enrichment; Figures 6E and 6F). Therefore, our findings are robust across different alignments and species backgrounds. Finally, a previous study identified 172 DNA damage response genes in human blood cells 24 hr after exposure to gamma irradiation (2 Gy) versus untreated controls (Ghandhi et al., 2015). We found that elephant ARs are not significantly enriched at these human DNA damage genes, nor are the ARs from the other species (data not shown), suggesting that elephant ARs (and orca ARs) are uniquely enriched at elephant DNA damage response genes.

To highlight one example, we identified elephant ARs associated with *CADM1* (cell adhesion molecule 1), a potent tumor suppressor (Murakami, 2005) that exhibits significantly decreased expression at 5 hr but increased expression at 24 hr in the elephant cells in response to DNA damage (Figure 6G, inset). One elephant AR associated with *CADM1* is located in a Vista-annotated enhancer conserved from marsupials to human, but changed in the elephant (Figure 6G). Overall, 34 conserved non-coding 50-bp regions associated with *CADM1* changed in the elephant lineage.

The human homologs of the elephant ARs at elephant DNA damage genes are significantly enriched for DNase-I-hypersensitive regions in 12 different human cell classes (Figures 6 and S1; *in silico* ChIP, ChIP-Atlas, FDR < 5% relative to random elements). These elements are also enriched for markers of putative functional elements in blood cells (H3K4me3, K3K4me2, and H3K4me1), for putative enhancers in cells derived from pluripotent stem cells (H3K27ac and EP300) (Figures 6 and S1), and for binding sites for 13 different TFs (Figures 6 and S1), including SOX2-binding sites, like the *VRK2-FANCL-BLC11A* locus detailed above. Thus, human homologs for elephant ARs at elephant DNA damage response genes are putative functional elements in human tissues, revealing candidates for shaping mutation and cancer resistance (Table S5).

Prediction of the Phenotypic Traits Impacted by Different ARs

In most cases, the phenotypes shaped by the different ARs identified in our study are unknown. To uncover candidate phenotypic effects, ARs were linked to hg19 protein-coding genes based on proximity using GREAT (McLean et al., 2010) (Table S6) and then cross-referenced with associated human phenotypes in the Human Phenotype Ontology (HPO) database (Köhler et al., 2017) (Table S1). We then tested for significant over-representations of ARs for specific human phenotypes relative to conserved regions (Table S7), uncovering human phenotypes significantly enriched for each of the six species' ARs (Figure 7A; FDR 5%). Over 80% of the phenotypes are species specific (Figure 7B), and we observed some concordance between the overt phenotypes of a particular species and the most significantly enriched clinical phenotypes for that species' ARs.

Among the top ten most significant enrichments for elephant ARs, we found an 11-fold enrichment for phenotypes involving aplasia/hypoplasia of the uvula (Figure 7B; FDR = 9×10^{-122}). This was not observed for any of the other five species (Figure 7C; Table S7). Consistent with this finding, elephants lack a uvula and evolved an unusually structured throat and short soft palate (Figure 7C) to suck water into their trunk and store large volumes of water in a pharyngeal pouch behind the tongue (Short, 1962). We also found a 10-fold enrichment for abnormalities of chromosome stability (FDR = 8×10^{-112}). Interestingly, both of these phenotypes are linked to *FANCL*, a hotspot for elephant ARs. Indeed, in addition to being a master regulator of DNA repair and chromosome stability (Moldovan and D'Andrea, 2009), the FA pathway impacts the development of some organ systems (OMIM 614083). Future studies are needed to determine which elephant ARs near *FANCL* are involved in DNA repair-related processes. Overall, we uncovered 279 different phenotypic traits linked to elephant ARs at various genes (Table S7).

Most of the top human phenotype enrichments for Hib bat ARs relate to anatomical abnormalities of the hands and feet, including synostosis involving metatarsal bones, syndactyly, clinodactyly, and other digit-related abnormalities (Figure 7D; Table S7). This finding is striking because wings are the bat's most conspicuous anatomical trait. Wings evolved from elongation of the metacarpals and phalanges and selection for webbed digits (Booker et al., 2016; Eckalbar et al., 2016; Sears et al., 2006). We also found a 22-fold enrichment for Stahl's ear, a genetic deformity that frequently creates pointy, bat-like ears in humans (Figure 7D). Interestingly, 37 genes with bat ARs are linked to syndactyly and digit phenotypes, but one gene, *SALL1*, is linked to both Stahl's ear and syndactyly, suggesting that some bat ARs at *SALL1* could shape ear morphology, while others shape digit development. Hib bat ARs are also implicated in other traits, including changes to vaginal and uterine development and cerebellar development (Table S7).

For the orca and dolphin, several top AR enrichments are related to eye development, including microcornea and abnormality of corneal size (Figure 7E; Table S7). Delphinidae have an unusual cornea that adapts the optics of the eye for vision in water (Fulton, 2014). Unique top enrichments for orca ARs are related to immunity and infection resistance phenotypes, such as a 23-fold enrichment for C8 deficiency (Figure 7E), which in humans influences susceptibility to some bacterial infections (see other enriched traits in Table S7).

The top human phenotype enrichment for mole rat ARs is glaucoma, an age-related degeneration of the retina and optic nerve (Figure 7F). This seems consistent with the mole rat's fossorial lifestyle, degenerated retina, and atrophied optic nerve (Crish et al., 2006). Indeed, other top mole rat AR enrichments included aplasia/hypoplasia affecting the anterior segment of the eye and aplasia/hypoplasia of the iris, uvea, and lens (Table S7). Different top enrichment examples are unossified vertebral bodies, absent in utero ossification of vertebral bodies, and absent in utero rib ossification (Figure 7F; Table S7). Interestingly, female queen mole rats will lengthen the lumbar vertebral column during puberty and pregnancy, which may function to increase body size to produce large litters (Dengler-Crish and Catania, 2009). Thus, mole rat ARs at genes linked to rib and vertebral body ossification may have roles in this unique phenotype.

Finally, the top squirrel AR enrichments are related to human pigmentation disorders, including a 23-fold enrichment for multiple lentigines (Leopard syndrome) and a 12-fold enrichment for partial albinism (Figure 7G; Table S7). This enrichment is concordant with the extraordinary patterns of squirrel fur pigmentation. Shared mechanisms are involved in skin and fur (and feather) pigmentation (Jimbow et al., 1976) (Figure 7G). Significant albeit more modest enrichments related to pigmentation were also observed for the Hib bat ARs, the other lineage in our study with pigmented fur. We found that many squirrel and Hib bat ARs linked to pigmentation abnormalities are enriched near *MITF* (microphthalmia-associated TF), which is linked to pigmentation and deafness disorders, melanoma, microphthalmia, and macrocephaly (George et al., 2016; Kawakami and Fisher, 2017). By comparing the squirrel ARs to DNase-seq datasets in the ChIP-Atlas for different human cell types, we found 10 *MITF* squirrel ARs that are located in epidermal melanocyte DNase-I-hypersensitive sites (Figures 7 and S1). These elements are therefore candidates for shaping pigmentation phenotypes in the squirrel and other mammalian species. Three of these ARs are binding sites for TFAP2C in human epidermal cells, indicating an upstream regulatory protein that acts on these elements (Figures 7 and S1). Our results provide testable predictions regarding the functional roles of specific ARs and putative links to human clinical phenotypes (Table S1).

DISCUSSION

Our study tested whether a comparative analysis of ARs in species with distinctive traits facilitates the discovery of candidate functional elements for the overt and clinically relevant traits exhibited by these species. From elephant, Hib bat, orca, dolphin, mole rat, and squirrel ARs, we identified a set of 33,283 candidate elements (7% of mammalian conserved regions tested). Multiple lines of evidence support the functionality of these elements, including selective constraint (conservation) from wallaby to human, regulatory protein binding in humans and mice, and evidence for accelerated evolution in specific lineages. Due to the elephant's distinctive body size and selective pressures to reduce somatic mutations and cancer risk, elephant ARs uniquely enriched at the *VRK2- FANCL-BCL11A* locus and at genes that respond to DNA damage in elephant blood cells are important candidate elements for shaping mammalian mutation and cancer resistance phenotypes. For the ARs uncovered in the other lineages, we found significant associations with human clinical phenotypes concordant with the species' overt phenotypes and show how discrete candidate elements for particular phenotypes, such as squirrel pigmentation, are revealed from contrasts with available human DNase-seq and/or ChIP-seq datasets. The results are expected to advance our understanding of the genetic basis of different mammalian phenotypes, including human clinical phenotypes.

The Identification of Candidate Genomic Elements for Preventing Somatic Mutations Could Reveal Mechanisms for Disease Prevention

Tumor risk is strongly influenced by cell division frequency and the associated DNA replication errors that cause somatic mutations (Vassilev and DePamphilis, 2017; Zhu et al., 2016). However, Peto's paradox notes that there is no association between body size and cancer risk among species and predicts that species with a massive body size evolved unique mechanisms to evade the increased risk of cancer (Caulin and Maley, 2011). One evolutionary strategy to evade cancer could be to reduce somatic mutation rates. Indeed, humans have a larger body size than mice and have a lower germline and somatic mutation rate than mice (Milholland et al., 2017). Compared to humans, the elephant is predicted to have a further 2.17-fold decrease in somatic mutation rate (Calabrese and Shibata, 2010). Importantly, while cancer risk is emphasized by Peto's paradox, somatic alterations such as mutations and copy-number variation (CNV) can influence the risk for many different maladies (McConnell et al., 2017; Poduri et al., 2013). For example, somatic mutations are an important risk factor in autism (Lim et al., 2017) and ~13%–41% of human frontal cortex neurons have at least one somatic CNV (McConnell et al., 2013). Widespread somatic variation has been identified in different human tissues (O'Huallachain et al., 2012). Thus, large species like the elephant presumably contend with a spectrum of increased health risks related to somatic mutations from cell division. Since the human homologs of the elephant ARs we found at the *FANCL* locus and at DNA damage response genes are conserved elements with apparent functional activity in humans, these elements may have functions that impact somatic mutation rates in humans (and mice).

Given growing evidence that somatic alterations are an important risk factor for neurological problems (Lim et al., 2017; McConnell et al., 2017; Poduri et al., 2013), it may be important that the human homologs for elephant ARs at the *VRK2-FANCL-BCL11A* locus, and at DNA damage response genes, are enriched for activity in neural cells. The elephant brain has 3–4 times more neurons than the human brain (Herculano-Houzel et al., 2014) and may have changed important regulatory elements to reduce the risk for somatic mutations in neural cell lineages. Our findings provide further support for Peto's paradox and substantially expand the landscape of candidate genomic elements for shaping elephant cancer resistance beyond *TP53* copy-number amplification (Abegglen et al., 2015; Sulak et al., 2016). Functional studies that build on our findings may involve engineering elephant genome sequences into human cell lines or transgenic mice to learn how specific elements shape DNA damage responses and somatic mutation rates. Deletions of the homologous elephant AR elements may also reveal effects on mutation rates and cancer, as well as studies of genetic or epigenetic variants impacting these elements in human populations.

Toward Evolutionary Phenotype Ontology for Conserved Noncoding Elements

The ARs in each of our six target species are significantly associated with specific and unique clinical human phenotypes. We observed apparent concordance between the top enriched human phenotypes and the overt traits in our target species. For example, bat ARs are highly and uniquely enriched for human phenotypes involving digit developmental abnormalities, and bats dramatically altered mammalian patterns of digit development to create a wing. In agreement with our findings, recent studies found bat ARs that are functional enhancers driving expression in digits, with putative roles in the development of the bat wing (Booker et al., 2016; Eckalbar et al., 2016). The categorization of functional genetic elements according to the phenotypes that they shape in different species is one important goal of the phylogenomics field (Hiller et al., 2012; Mabee et al., 2007; Manda et al., 2015). Identifying phenotypic equivalence (phenologs) across species is also thought to be important for modeling human disease processes in other species (McGary et al., 2010; Robinson and Webber, 2014). Most approaches toward these goals have focused on protein-coding sequence changes. However, changes in regulatory regions frequently play the major roles in the evolution of new phenotypes (Carroll, 2008; Wray, 2007). A disproportionate number of the elements we uncovered and linked to phenotypes are noncoding, providing an initial framework for an evolutionary phenotype ontology for conserved noncoding elements. Targeted loss-of-function or sequence replacement studies in transgenic mice or human cell lines (Yoshimi et al., 2016) are expected to elucidate how the different functional classes of ARs we uncovered in different species (e.g., enhancers, promoters, and insulators) function and shape particular traits.

EXPERIMENTAL PROCEDURES

Elephant Blood Samples

Whole-blood samples from an adult female African elephant supplied by Utah's Hogle Zoo (Salt Lake City, UT) were used for RNA-sequencing experiments. All experimental procedures were reviewed and approved by the Hogle Zoo's ethical and scientific review board.

Identification of ARs in the Target Species

Our study is based on the species available in the hg19/GRCh37 vertebrate 100-way UCSC multiple genome alignment file (MAF). From the background mammalian species, we identified 50-bp conserved regions using PhastCons (Hubisz et al., 2011; Siepel et al., 2005) with the following parameters: expected. length = 45, target.coverage = 0.3, rho = 0.31). Next, ARs for each of the six target species were defined in the conserved regions using RPHAST (Hubisz et al., 2011; Pollard et al., 2010). Statistically significant ARs were defined with a FDR threshold of 5%. Full details can be found in Supplemental Experimental Procedures.

Analysis of GC-Biased Gene Conversion versus Adaptive Nucleotide Substitutions in Elephant ARs

We used previously published methods (Kostka et al., 2012) to test for biased gene conversion versus adaptive patterns of nucleotide substitutions at elephant ARs in the VRK2-FANCL-BCL11A hotspot (Supplemental Experimental Procedures).

Identification of ARs in Human and Mouse Putative Functional Genomic Elements

To test whether ARs are located in putative functional elements in the human and mouse genomes and characterize the biochemical activity of these elements, we used publically available DNase-seq and ChIP-seq datasets in the ChIP-Atlas (Supplemental Experimental Procedures).

Transcriptome Profiling in Primary Elephant Peripheral Blood Lymphocytes

Whole-blood samples from a female African elephant supplied by Utah's Hogle Zoo were used for RNA-sequencing experiments. Processing of blood for live cell experiments began within 1 hr of the blood draw. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Paque density gradient followed by three washes in large volumes of PBS to remove platelets. Wholetranscriptome profiling was performed using the Illumina TruSeq Stranded RNA kit with Ribo-Zero Gold. DESeq2 (Anders et al., 2013) was used to calculate FPKM (fragments per kilobase per million reads) for annotated elephant genes (GenBank: GCA_000001905.1) and define differentially expressed genes. Differentially expressed genes are identified by comparing 1, 5, and 24 hr after radiation to time 0 or 1, 5, and 24 hr without radiation, respectively. Statistically significant differentially expressed genes were defined at an FDR threshold of 5%.

Statistical Analysis of AR Enrichments at DNA Damage Genes

ARs and mammalian conserved regions (CRs) were assigned to genes based on custom code adopting the proximity algorithm approach of GREAT (McLean et al., 2010). The LOLA (Locus Overlap Analysis) package in R was then used to compute statistically significant AR enrichments at specific gene classes, such as 5-hr and 24-hr elephant DNA damage response genes, relative to the background of conserved regions. Significant enrichment effects are $p < 0.05$.

DATA AND SOFTWARE AVAILABILITY

The accession number for the elephant primary blood cell RNA-seq datasets reported in this paper is GEO: GSE107117.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Hilary Coon, Jared Rutter, Nicki Camp, Suzi Mansour, Monica Vetter, and members of the Gregg lab for reviewing earlier versions of the manuscript. We thank Dr. Tim Mosbruger for help with the analysis of the elephant RNA-seq data, Dr. Melissa J. Hubisz for help with RPHAST, and Dr. Glenn Bartlett for Stahl Ear photos. Research reported in this publication utilized the High-Throughput Genomics and Bioinformatic Analysis Shared Resource at the Huntsman Cancer Institute supported by award P30CA042014 from the National Cancer Institute. This work was supported by the NIH (grant R01MH109577 to C.G.). The elephant genomic work was supported by the Ringling Bros. Children's Fund of The Feld Family Foundation, Li-Fraumeni Syndrome Association, Kneaders Bakery & Cafe (Hope Fights Childhood Cancer Campaign), and the Soccer For Hope Foundation. J.D.S. is supported by an Edward B. Clark, MD Chair in Pediatric Research. C.G. is a New York Stem Cell Foundation Robertson Neuroscience Investigator. This research was supported by the New York Stem Cell Foundation.

References

Abdelalim EM, Emara MM, Kolatkar PR. 2014; The SOX transcription factors as key players in pluripotent stem cells. Stem Cells Dev. 23:2687–2699. [PubMed: 25127330]

Abegglen LM, Caulin AF, Chan A, Lee K, Robinson R, Campbell MS, Kiso WK, Schmitt DL, Waddell PJ, Bhaskara S, et al. 2015; Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. JAMA. 314:1850–1860. [PubMed: 26447779]

Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD. 2013; Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc. 8:1765–1786. [PubMed: 23975260]

Bird CP, Stranger BE, Liu M, Thomas DJ, Ingle CE, Beazley C, Miller W, Hurles ME, Dermitzakis ET. 2007; Fast-evolving noncoding sequences in the human genome. Genome Biol. 8:R118. [PubMed: 17578567]

Blanco S, Klimcakova L, Vega FM, Lazo PA. 2006; The subcellular localization of vaccinia-related kinase-2 (VRK2) isoforms determines their different effect on p53 stability in tumour cell lines. FEBS J. 273:2487–2504. [PubMed: 16704422]

Booker BM, Friedrich T, Mason MK, VanderMeer JE, Zhao J, Eckalbar WL, Logan M, Illing N, Pollard KS, Ahituv N. 2016; Bat accelerated regions identify a bat forelimb specific enhancer in the HoxD locus. PLoS Genet. 12:e1005738–e21. [PubMed: 27019019]

Boyd JL, Skove SL, Rouanet JP, Pilaz L-J, Bepler T, Gordân R, Wray GA, Silver DL. 2015; Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. Curr Biol. 25:772–779. [PubMed: 25702574]

Calabrese P, Shibata D. 2010; A simple algebraic cancer equation: calculating how cancers may arise with normal mutation rates. BMC Cancer. 10:3. [PubMed: 20051132]

Capra JA, Erwin GD, McKinsey G, Rubenstein JLR, Pollard KS. 2013; Many human accelerated regions are developmental enhancers. Philos Trans R Soc Lond B Biol Sci. 368:20130025–20130025. [PubMed: 24218637]

Carey HV, Andrews MT, Martin SL. 2003; Mammalian hibernation: cellular and molecular responses to depressed metabolism and low temperature. Physiol Rev. 83:1153–1181. [PubMed: 14506303]

Carroll SB. 2008; Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 134:25–36. [PubMed: 18614008]

Caulin AF, Maley CC. 2011; Peto's paradox: evolution's prescription for cancer prevention. Trends Ecol Evol. 26:175–182. [PubMed: 21296451]

Crish SD, Dengler-Crish CM, Catania KC. 2006; Central visual system of the naked mole-rat (Heterocephalus glaber). Anat Rec A Discov Mol Cell Evol Biol. 288:205–212. [PubMed: 16419086]

Dengler-Crish CM, Catania KC. 2009; Cessation of reproductionrelated spine elongation after multiple breeding cycles in female naked mole-rats. Anat Rec (Hoboken). 292:131–137. [PubMed: 18951517]

Eckalbar WL, Schlebusch SA, Mason MK, Gill Z, Parker AV, Booker BM, Nishizaki S, Muswamba-Nday C, Terhune E, Nevonen KA, et al. 2016; Transcriptomic and epigenomic characterization of the developing bat wing. Nat Genet. 48:528–536. [PubMed: 27019111]

ENCODE Project Consortium. 2012; An integrated encyclopedia of DNA elements in the human genome. Nature. 489:57–74. [PubMed: 22955616]

Fulton, JT. Vision, hearing and language of the dolphin, *Tursiops trancatus*. 2014. http://neuronresearch.net/dolphin/index.html

George A, Zand DJ, Hufnagel RB, Sharma R, Sergeev YV, Legare JM, Rice GM, Scott Schwoerer JA, Rius M, Tetri L, et al. 2016; Biallelic mutations in MITF cause coloboma, osteopetrosis, microphthalmia, macrocephaly, albinism, and deafness. Am J Hum Genet. 99:1388–1394. [PubMed: 27889061]

Ghandhi SA, Smilenov LB, Elliston CD, Chowdhury M, Amundson SA. 2015; Radiation dose-rate effects on gene expression for human biodosimetry. BMC Med Genomics. 8:22. [PubMed: 25963628]

Harmston N, Ing-Simmons E, Tan G, Perry M, Merkenschlager M, Lenhard B. 2017; Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. Nat Commun. 8:441. [PubMed: 28874668]

Herculano-Houzel S, Avelino-de-Souza K, Neves K, Porfírio J, Messeder D, Mattos Feijó L, Maldonado J, Manger PR. 2014; The elephant brain in numbers. Front Neuroanat. 8:46. [PubMed: 24971054]

Hiller M, Schaar BT, Indjeian VB, Kingsley DM, Hagey LR, Bejerano G. 2012; A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. Cell Rep. 2:817–823. [PubMed: 23022484]

Hubisz MJ, Pollard KS. 2014; Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. Curr Opin Genet Dev. 29:15–21. [PubMed: 25156517]

Hubisz MJ, Pollard KS, Siepel A. 2011; PHAST and RPHAST: phylogenetic analysis with space/time models. Brief Bioinform. 12:41–51. [PubMed: 21278375]

Jimbow K, Quevedo WC Jr, Fitzpatrick TB, Szabo G. 1976; Some aspects of melanin biology: 1950-1975. J Invest Dermatol. 67:72–89. [PubMed: 819593]

Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, et al. 2013; DNA-binding specificities of human transcription factors. Cell. 152:327–339. [PubMed: 23332764]

Kaltcheva MM, Lewandoski M. 2016; Evolution: enhanced footing for snake limb development. Curr Biol. 26:R1237–R1240. [PubMed: 27923134]

Kawakami A, Fisher DE. 2017; The master role of microphthalmia-associated transcription factor in melanocyte and melanoma biology. Lab Invest. 97:649–656. [PubMed: 28263292]

Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, Ward LD, Birney E, Crawford GE, Dekker J, et al. 2014; Defining functional DNA elements in the human genome. Proc Natl Acad Sci USA. 111:6131–6138. [PubMed: 24753594]

Khaled WT, Lee SC, Stingl J, Chen X, Ali HR, Rueda OM, Hadi F, Wang J, Yu Y, Chin S-F, et al. 2015; (1AD). BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. Nat Commun. 6:1–10.

Kim SY, Pritchard JK. 2007; Adaptive evolution of conserved noncoding elements in mammals. PLoS Genet. 3:1572–1586. [PubMed: 17845075]

Klerkx EPF, Lazo PA, Askjaer P. 2009; Emerging biological functions of the vaccinia-related kinase (VRK) family. Histol Histopathol. 24:749–759. [PubMed: 19337973]

Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, et al. 2017; The human phenotype ontology in 2017. Nucleic Acids Res. 45(D1):D865–D876. [PubMed: 27899602]

Kostka D, Hubisz MJ, Siepel A, Pollard KS. 2012; The role of GC-biased gene conversion in shaping the fastest evolving regions of the human genome. Mol Biol Evol. 29:1047–1057. [PubMed: 22075116]

Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, Tissiéres V, Pickle CS, Plajzer-Frick I, Lee EA, et al. 2016; Progressive loss of function in a limb enhancer during snake evolution. Cell. 167:633–642.e11. [PubMed: 27768887]

Lim ET, Uddin M, De Rubeis S, Chan Y, Kamumbu AS, Zhang X, D'Gama AM, Kim SN, Hill RS, Goldberg AP, et al. Autism Sequencing Consortium. 2017; Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. Nat Neurosci. 20:1217–1224. [PubMed: 28714951]

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, et al. Broad Institute Sequencing Platform and Whole Genome Assembly Team; Baylor College of Medicine Human Genome Sequencing Center Sequencing Team; Genome Institute at Washington University. 2011; A high-resolution map of human evolutionary constraint using 29 mammals. Nature. 478:476–482. [PubMed: 21993624]

Mabee PM, Ashburner M, Cronk Q, Gkoutos GV, Haendel M, Segerdell E, Mungall C, Westerfield M. 2007; Phenotype ontologies: the bridge between genomics and evolution. Trends Ecol Evol. 22:345–350. [PubMed: 17416439]

Manda P, Balhoff JP, Lapp H, Mabee P, Vision TJ. 2015; Using the phenoscape knowledgebase to relate genetic perturbations to phenotypic evolution. Genesis. 53:561–571. [PubMed: 26220875]

McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, Cowing-Zitron C, Shumilina S, Lasken RS, Vermeesch JR, Hall IM, Gage FH. 2013; Mosaic copy number variation in human neurons. Science. 342:632–637. [PubMed: 24179226]

McConnell MJ, Moran JV, Abyzov A, Akbarian S, Bae T, Cortes-Ciriano I, Erwin JA, Fasching L, Flasch DA, Freed D, et al. 2017; Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. Science. 356:eaal1641–11. [PubMed: 28450582]

McGary KL, Park TJ, Woods JO, Cha HJ, Wallingford JB, Marcotte EM. 2010; Systematic discovery of nonobvious human disease models through orthologous phenotypes. Proc Natl Acad Sci USA. 107:6544–6549. [PubMed: 20308572]

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010; GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 28:495–501. [PubMed: 20436461]

McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X, Menke DB, Schaar BT, et al. 2011; Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature. 471:216–219. [PubMed: 21390129]

Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. 2017; Differences between germline and somatic mutation rates in humans and mice. Nat Commun. 8:15183. [PubMed: 28485371]

Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, Ebbert A, Riley ZL, Royall JJ, Aiona K, et al. 2014; Transcriptional landscape of the prenatal human brain. Nature. 508:199–206. [PubMed: 24695229]

Moldovan G-L, D'Andrea AD. 2009; How the Fanconi anemia pathway guards the genome. Annu Rev Genet. 43:223–249. [PubMed: 19686080]

Murakami Y. 2005; Involvement of a cell adhesion molecule, TSLC1/IGSF4, in human oncogenesis. Cancer Sci. 96:543–552. [PubMed: 16128739]

O'Huallachain M, Karczewski KJ, Weissman SM, Urban AE, Snyder MP. 2012; Extensive genetic variation in somatic human tissues. Proc Natl Acad Sci USA. 109:18018–18023. [PubMed: 23043118]

Poduri A, Evrony GD, Cai X, Walsh CA. 2013; Somatic mutation, genomic variation, and neurological disease. Science. 341:1237758. [PubMed: 23828942]

Pollard KS, Salama SR, King B, Kern AD, Dreszer T, Katzman S, Siepel A, Pedersen JS, Bejerano G, Baertsch R, et al. 2006a; Forces shaping the fastest evolving regions in the human genome. PLoS Genet. 2:e168. [PubMed: 17040131]

Pollard KS, Salama SR, Lambert N, Lambot M-A, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, Siepel A, et al. 2006b; An RNA gene expressed during cortical development evolved rapidly in humans. Nature. 443:167–172. [PubMed: 16915236]

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010; Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 20:110–121. [PubMed: 19858363]

Prabhakar S, Noonan JP, Pääbo S, Rubin EM. 2006; Accelerated evolution of conserved noncoding sequences in humans. Science. 314:786. [PubMed: 17082449]

Prabhakar S, Visel A, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Morrison H, Fitzpatrick DR, Afzal V, et al. 2008; Human-specific gain of function in a developmental enhancer. Science. 321:1346–1350. [PubMed: 18772437]

Robinson PN, Webber C. 2014; Phenotype ontologies and crossspecies analysis for translational research. PLoS Genet. 10:e1004268. [PubMed: 24699242]

Sears KE, Behringer RR, Rasweiler JJ 4th, Niswander LA. 2006; Development of bat flight: morphologic and molecular evolution of bat wing digits. Proc Natl Acad Sci USA. 103:6581–6586. [PubMed: 16618938]

Short RV. 1962; The peculiar lungs of the elephant. New Sci. 16:570–572.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005; Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 15:1034–1050. [PubMed: 16024819]

Sulak M, Fong L, Mika K, Chigurupati S, Yon L, Mongan NP, Emes RD, Lynch VJ. 2016; TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. eLife. 5:1850.

Tollis M, Schiffman JD, Boddy AM. 2017; Evolution of cancer suppression as revealed by mammalian comparative genomics. Curr Opin Genet Dev. 42:40–47. [PubMed: 28161621]

Vassilev A, DePamphilis ML. 2017; Links between DNA replication, stem cells and cancer. Genes (Basel). 8:E45. [PubMed: 28125050]

Wray GA. 2007; The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 8:206–216. [PubMed: 17304246]

Yoshimi K, Kunihiro Y, Kaneko T, Nagahora H, Voigt B, Mashimo T. 2016; ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. Nat Commun. 7:10431. [PubMed: 26786405]

Yu Y, Wang J, Khaled W, Burke S, Li P, Chen X, Yang W, Jenkins NA, Copeland NG, Zhang S, Liu P. 2012; Bcl11a is essential for lymphoid development and negatively regulates p53. J Exp Med. 209:2467–2483. [PubMed: 23230003]

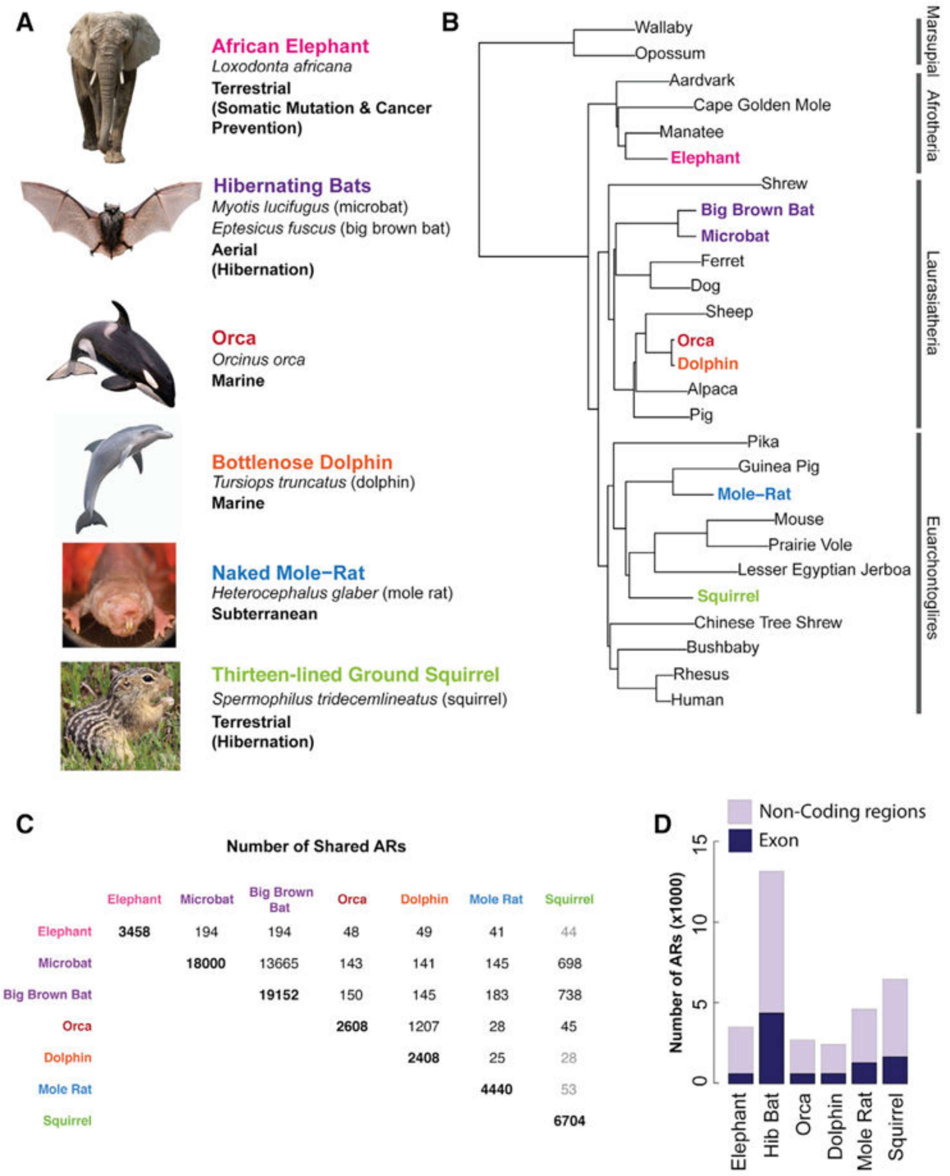
Zhu L, Finkelstein D, Gao C, Shi L, Wang Y, López-Terrada D, Wang K, Utley S, Pounds S, Neale G, et al. 2016; Multi-organ mapping of cancer risk. Cell. 166:1132–1146.e7. [PubMed: 27565343]

Highlights

- Accelerated evolution in terrestrial, aerial, marine, and subterranean mammals
- Accelerated regions (ARs) uncover diverse putative functional elements
- Elephant ARs reveal candidate mechanisms to decrease mutations and cancer risk
- Human AR homologs indicate concordance between clinical and species' phenotypes

In Brief

Ferris et al. report an analysis of accelerated evolution in the elephant, little brown bat, big brown bat, orca, dolphin, naked mole rate, and thirteen-lined ground squirrel that reveals candidate functional genomic elements for shaping somatic mutation rate, cancer risk, digit development, immunity, glaucoma, pigmentation, and other clinical phenotypes.



Number of Shared ARs

	Elephant	Microbat	Big Brown Bat	Orca	Dolphin	Mole Rat	Squirrel
Elephant	**3458**	194	194	48	49	41	44
Microbat		**18000**	13665	143	141	145	698
Big Brown Bat			**19152**	150	145	183	738
Orca				**2608**	1207	28	45
Dolphin					**2408**	25	28
Mole Rat						**4440**	53
Squirrel							**6704**

Figure 1. Identification of ARs in the Elephant, Hibernating Bat, Orca, Dolphin, Mole Rat, and Squirrel in a Shared Background of Mammalian Conserved Elements

(A) Images of the six target lineages, indicating habitats, and some distinctive traits.

(B) Phylogenetic tree of the species in hg19-rooted multiple genome alignment study. Target species are indicated in colored text. Background species for the identification of conserved genomic elements are black.

(C) Table shows the number of ARs that are in common between the seven target species. Bold text indicates the number of ARs for one species. Black text indicates shared ARs between two species where the overlap is greater than chance ($p < 0.05$, hypergeometric test relative to 660,851 conserved elements). Gray text indicates shared ARs where the overlap is not statistically significant.

(D) The number of ARs identified for each of the target species in hg19 annotated exons and noncoding elements.

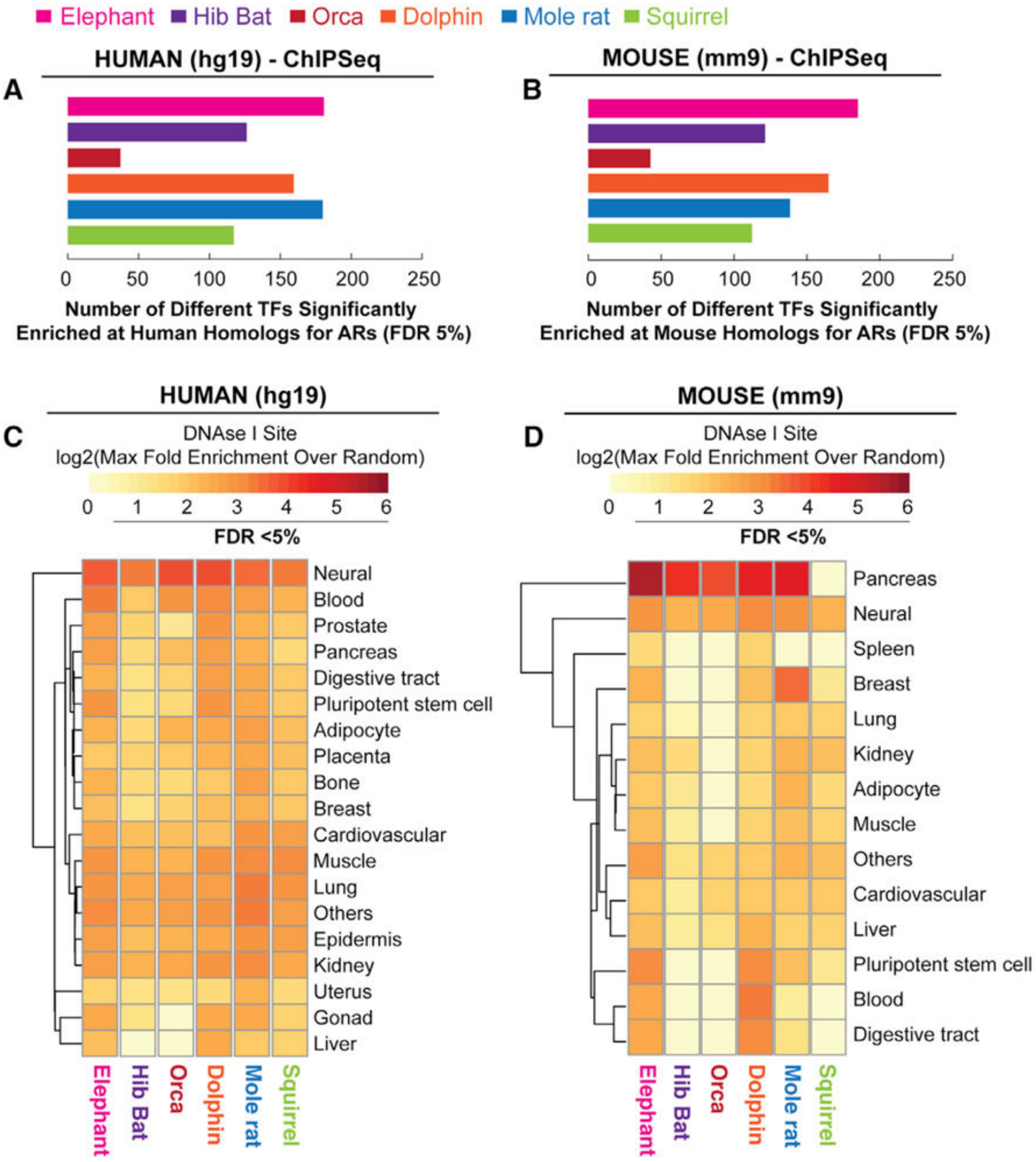


Figure 2. Human and Mouse Homologs for ARs Are Enriched for Transcription-Factor-Binding Sites and Are Active Elements in Diverse Tissues and Cell Types

(A and B) The bar graphs show the number of different TFs significantly enriched for binding sites in the human (A) and mouse (B) homologs the species' ARs (FDR 5% or less in at least one cell type relative to random elements; *in silico* ChIP in ChIP-Atlas).

(C and D) Heatmaps show enrichments for DNase- I-hypersensitive sites in the homologous elements for the species' ARs in various human (C) and mouse (D) cell types (*in silico* ChIP). All shown enrichments greater than zero are significant (FDR < 5%).

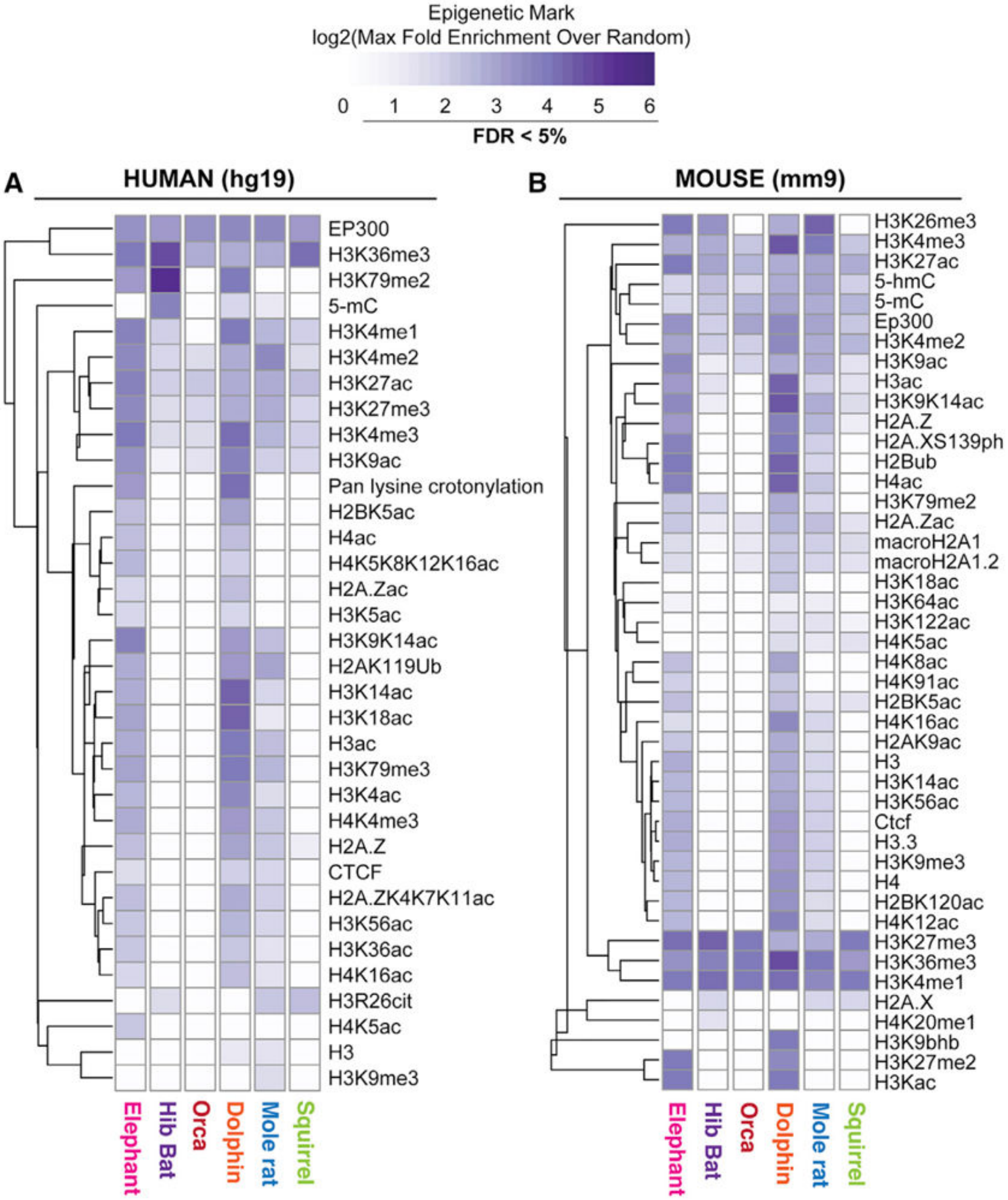


Figure 3. ARs from Different Species Are Biochemically Active Elements in Humans and Mice and Are Differentially Enriched for Specific Epigenetic Marks

(A and B) Heatmaps show the enrichment for different biochemical marks in the homologous elements for the species' ARs in human (A) and mouse (B) ChIP experiments. All enrichments greater than zero (white squares) are statistically significant compared to random elements (FDR < 5%, *in silico* ChIP, ChIP-Atlas).

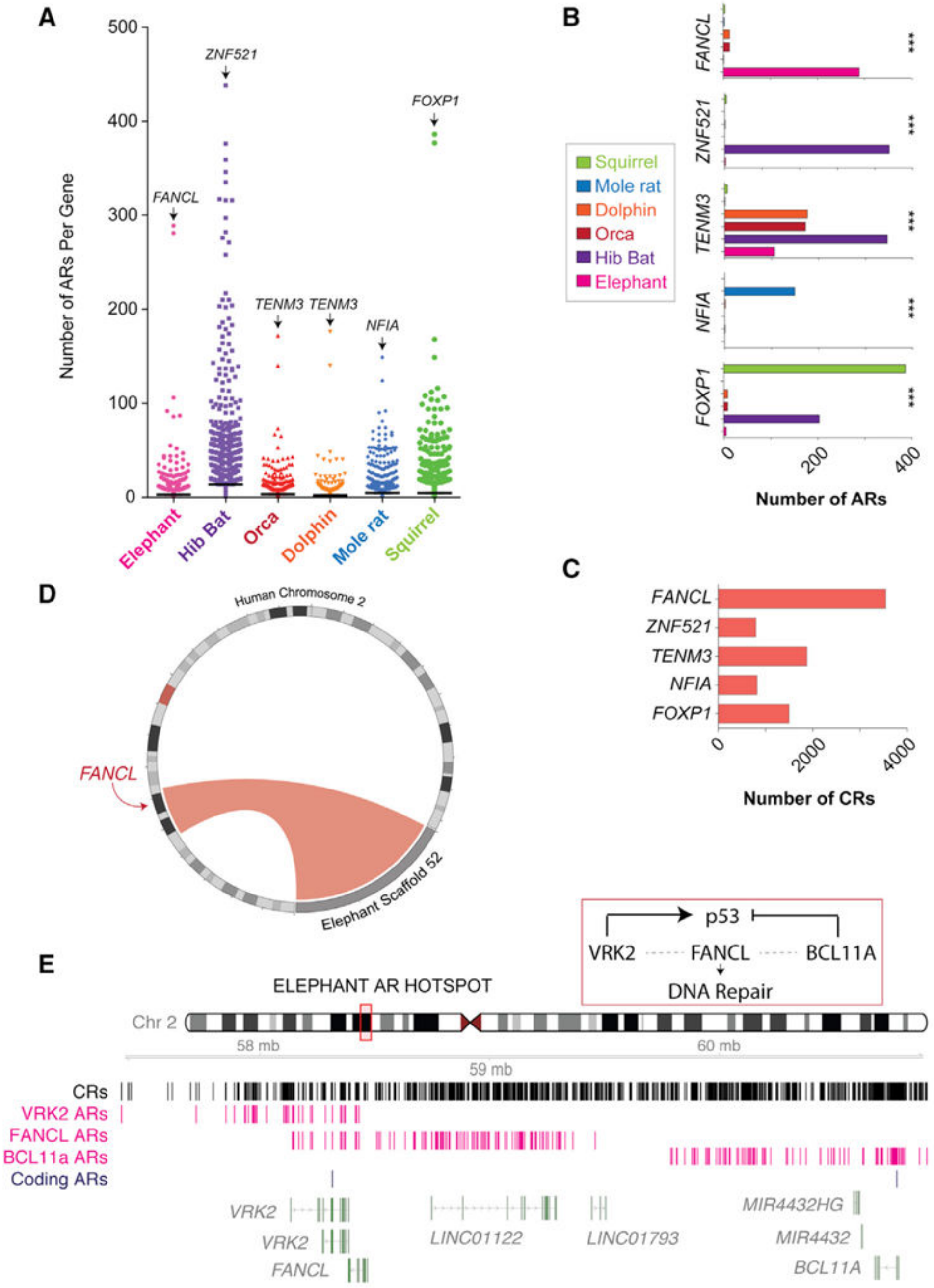


Figure 4. The Top Hotspot for Elephant ARs Is Associated with the Gene FANCL, a Master Regulator of DNA Repair

(A) Plot of the number of ARs per gene for each of the target species (genes without ARs are not shown). A subpopulation of genes is highly enriched for ARs in each species, revealing hotspots. *FANCL*, a master regulator of DNA repair, is associated with the largest number of ARs in the elephant genome.

(B) Bar charts show the number of ARs for genes that are top AR hotspots. *FANCL, ZNF521*, and *NFIA* are highly species-specific AR hotspots, while *TENM3* and *FOXP1* are hotspots for multiple species' ARs. Significant lineage-specific AR enrichment effects were determined with a twotailed chi-square comparing the observed number of ARs in each lineage versus the expected number based on the total ARs in each lineage. Significant lineage-dependent AR enrichment effects were observed for all 5 hotspots (***p < 0.0001).

(C) Bar chart shows the number of mammalian conserved regions associated with each of the AR hotspot genes in (B).

(D) Circle plot shows that the synteny of the human genomic region harboring *FANCL* is conserved in the elephant and corresponds to a region on scaffold 52 of the elephant genome assembly.

(E) Human genome tracks show the location of mammalian conserved regions (CRs, black) and elephant ARs (pink) across the human *VRK2-FANCL-BCL11A* region. Tracks of the ARs assigned to *VRK2, FANCL*, and *BCL11A* (pink) are shown separately and some ARs are assigned to both *VRK2* and *FANCL* (overlapping tracks). ARs also are associated with the lncRNAs, *LINC01122*, and *LINC01793* and the microRNAs *MIR4432HG* and *MIR4432*, which have unknown functions. The two ARs that overlap with exons are shown in blue (coding ARs). The inset schematic summarizes known interactions between p53 and the genes in the *VRK2-FANCL-BCL11A* region.

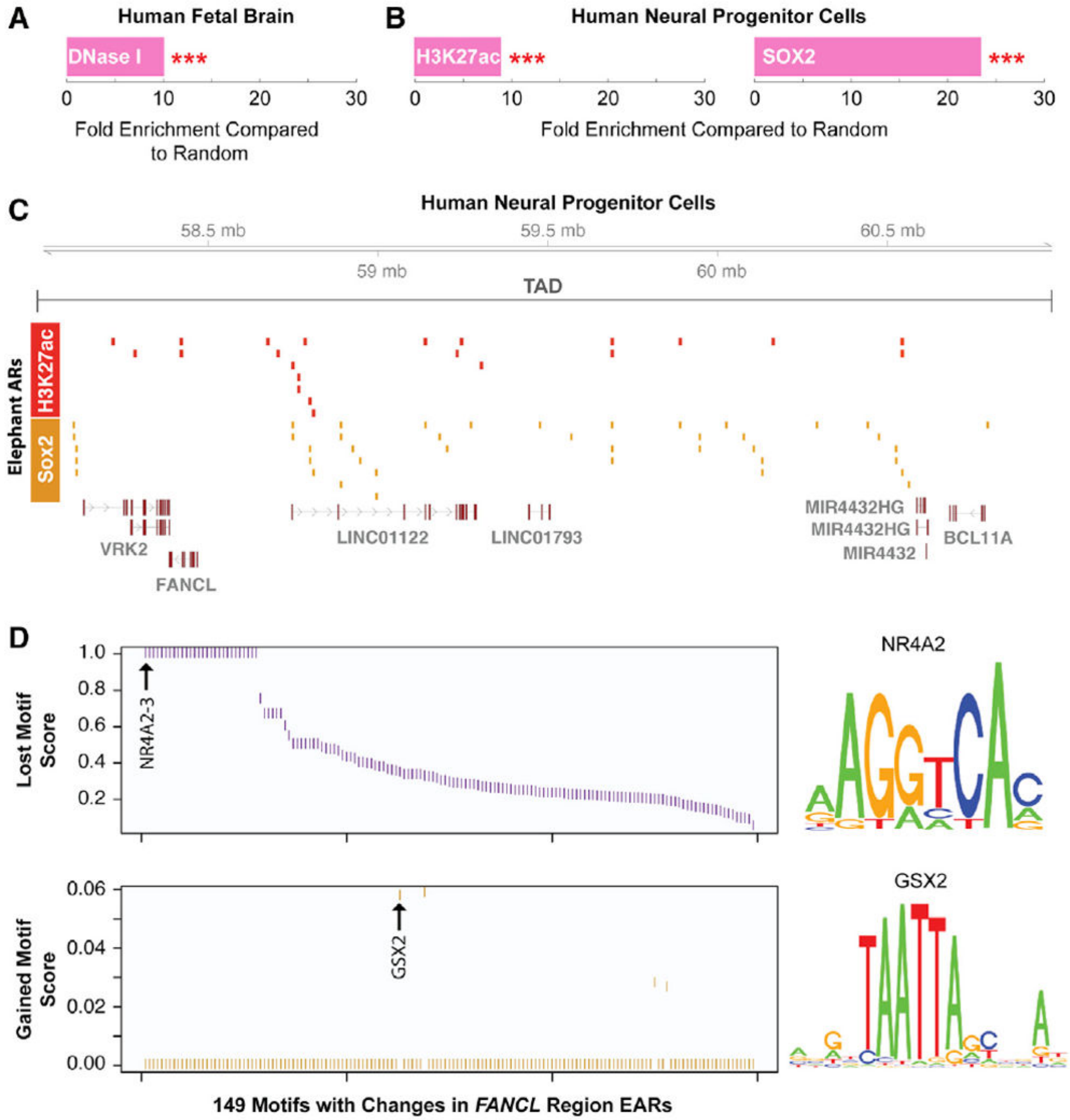


Figure 5. Elephant ARs at the *VRK2-FANCL-BCL11A* Locus Are Biochemically Active in Dividing Neural Cells, and the Elephant Changed the Conserved Regulatory Architecture in These Elements Compared to Smaller Mammals

(A and B) Bar plots show the fold enrichment effect for *VRK2-FANCL-BCL11A* locus elephant ARs in human fetal brain DNase-I-hypersensitive regions (A), and H3K27ac+ peaks and SOX2-binding sites in human neural progenitor cells (B). ***FDR < 0.0001 compared to 100 permutations of randomly sampled genomic regions.

(C) Genome tracks at the *VRK2-FANCL-BCL11A* locus for elephant ARs located in human neural progenitor H3K27ac ChIP-seq peaks (red) and SOX2B-binding site peaks (orange) (peak call $q < 1 \times 10^{-5}$). A topologically associating domain (TAD) encompasses the *VRK2-FANCL-BCL11A* locus in humans (dark gray bar).

(D) Dot plots show scores for conserved TFBS motifs that were lost (top plot, lost motif score) and new TFBS motifs that were gained (bottom plot, gained motif score) in the elephant genome at *VRK2-FANCL-BCL11A* elephant ARs compared to the homologous manatee, dog, and human sequences. While 709 TFBS motifs were tested, the plots depict the 149 motifs that changed (lost or gained sites). The elephant lost a relatively large number of TFBS motifs that are conserved in the manatee, dog, and human, such as PROX1-binding site motifs (top plot). The elephant gained new TFBS motifs, such as the POU3F3 (bottom plot) motif, compared to the manatee, dog, and human.

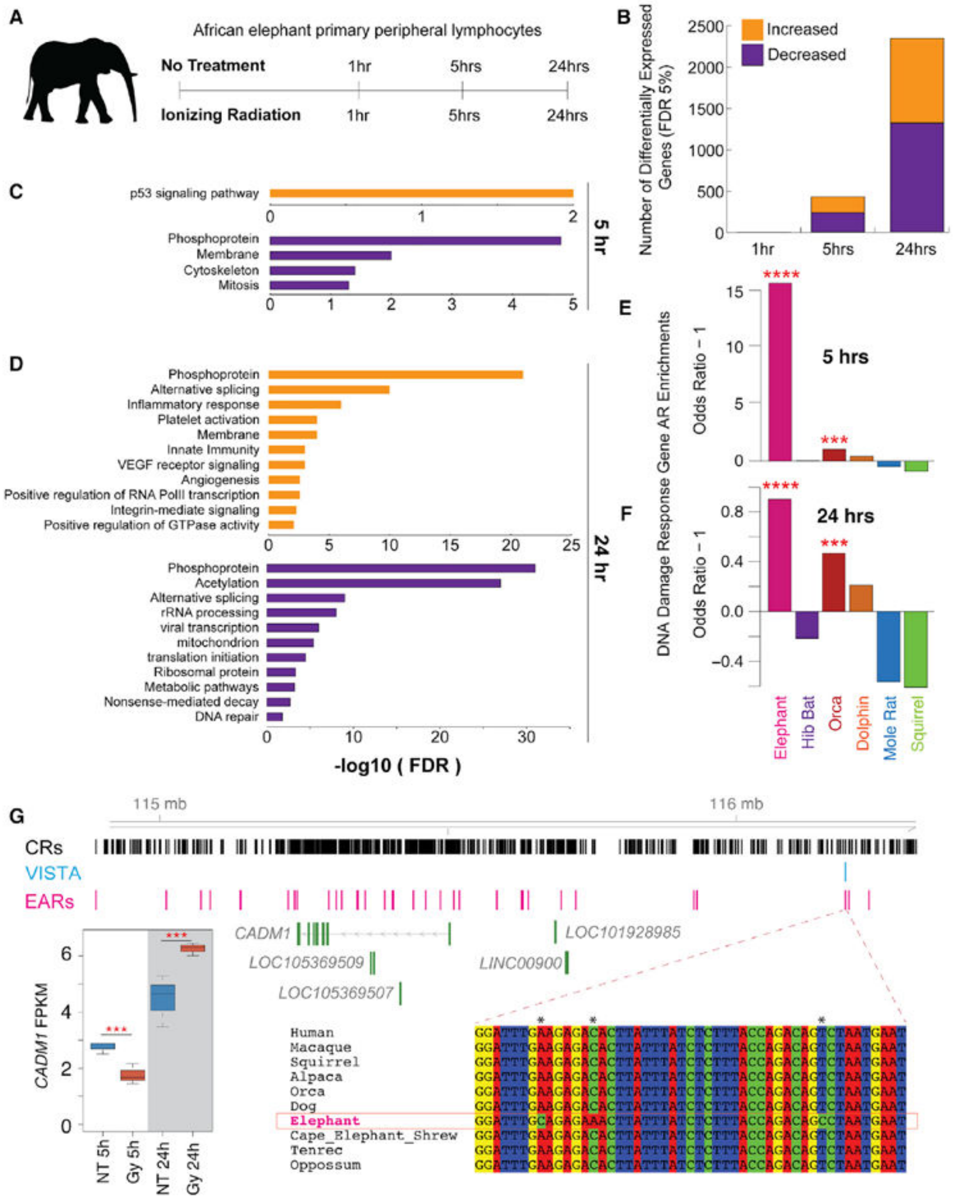


Figure 6. Genes Involved in the Intrinsic Elephant DNA Damage Response Program Are Significantly and Uniquely Enriched for Elephant ARs

(A) Experimental design to uncover a DNA damage response gene expression program in primary elephant peripheral lymphocytes. Cells are acutely irradiated (Gy2.0) or not (no treatment controls) at time 0 and then harvested for RNA and transcriptome profiling at 1, 5, and 24 hr.

(B) Number of genes with significantly increased (orange) or decreased (purple) expression in irradiated versus control cells at each time point (5% FDR).

(C and D) Gene ontology enrichment analysis for the 5-hr (C) and 24-hr (D) genes with significantly increased (orange bars) or decreased (purple) expression. The bars indicate the −log10 of the FDR and all bars are statistically significant for each term. Distinct and diverse functional categories of genes change their expression at the 5-hr and 24-hr time points.

(E and F) The bar charts show that genes differentially expressed at the 5-hr (E) and 24-hr (F) time points are significantly enriched for elephant ARs relative to the background of conserved elements. The odds ratio for an AR to be in a conserved element associated with an elephant DNA damage response gene is shown. Only elephant and orca ARs are significantly enriched at 5-hr and 24-hr elephant DNA damage response genes and the odds ratio for elephant ARs is several fold higher than the ARs for any of the other species. ****$p < 1 \times 10^{-10}$, ***$p < 0.0001$.

(G) Human genome tracks showing the location of conserved regions (CRs, black), Vista-annotated enhancers (VISTA, light blue), and elephant ARs (EARs, pink) at the *CADM1* locus, a DNA damage response gene and tumor suppressor. *CADM1* exhibits significantly decreased expression at the 5-hr time point and increased expression at the 24-hr time point following acute irradiation relative to control cells (inset boxplot, n = 3, FPKM, fragments per kilobase per million reads). The multiple sequence alignment for a *CADM1* VISTA annotated enhancer that overlaps with an elephant AR shows conservation from opossum to human, but specific nucleotide changes in the elephant (marked by asterisk).

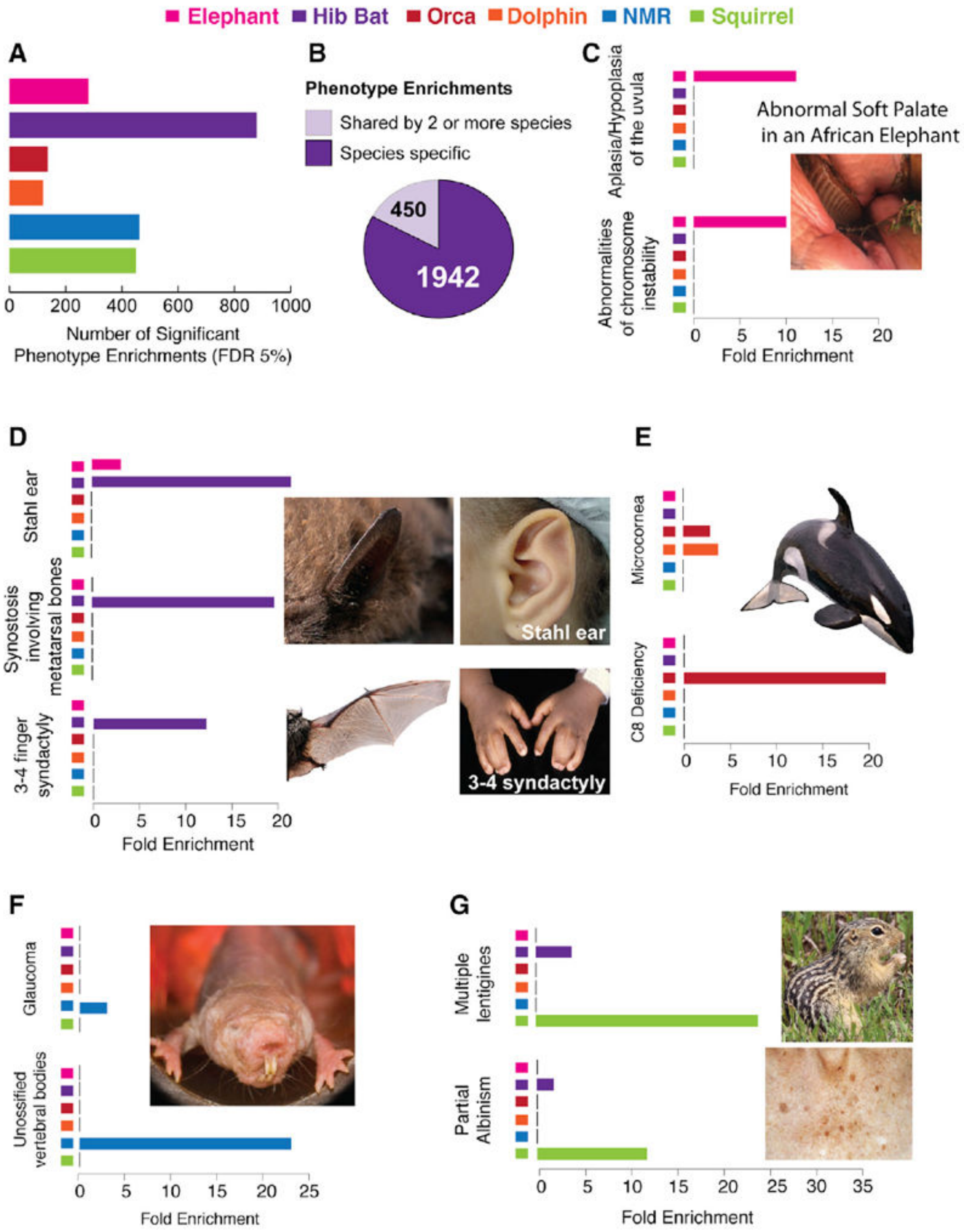


Figure 7. Prediction of Phenotypes Impacted by Species' ARs Based on Human Clinical Genetics Data

(A) Bar graph shows the number of statistically significant human phenotype enrichments for the ARs from each species compared to the background of conserved elements (FDR 5%).

(B) Pie chart shows the number of significant phenotype enrichments that are shared by two or more target species versus those that are species-specific.

(C–G) The bar charts show the fold enrichment for top significant phenotypic enrichments for each species' ARs compared to the background of conserved regions (FDR < 1 × 10^{-100} in all cases shown; Table S7). (C) Elephant ARs are significantly and uniquely enriched for phenotypes associated with hypoplasia of the uvula or abnormal chromosome instability. The photo shows the back of the throat of an adult female African elephant, revealing the absence of a uvula and a unique soft palate, consistent with predictions from the ARs. (D) Hib bat ARs are significantly associated with Stahl’s ear, and digit abnormalities, including metacarpal and metatarsal synostosis and syndactyly. Photos show morphological comparisons of human Stahl ear and the bat ear, and human syndactyly phenotypes and the bat wing. (E) Top orca AR phenotype enrichments are related to corneal developmental phenotypes (microcornea) and immunity mechanisms (C8 deficiency). (F) Top naked mole rat (NMR) AR enrichments include human phenotypes involving glaucoma and unossified vertebral bodies. (G) Top enrichments for squirrel ARs are related to pigmentation abnormalities, including multiple lentigines and partial albinism. Photos show the distinctive fur pigmentation pattern in the squirrel and examples of human multiple lentigines.