Title

Deciphering the role of structural variation in human evolution: a functional perspective

Authors

Karageorgiou, Charikleia
Gokcumen, Omer
Dennis, Megan Y

Publication Date

2024-10-01

DOI

10.1016/j.gde.2024.102240

Peer reviewed

HHS Public Access

Author manuscript

Curr Opin Genet Dev. Author manuscript; available in PMC 2024 October 17.

Published in final edited form as:

Curr Opin Genet Dev. 2024 October ; 88: 102240. doi:10.1016/j.gde.2024.102240.

Deciphering the role of structural variation in human evolution: a functional perspective

Charikleia Karageorgiou[1,*], **Omer Gokcumen**[1], **Megan Y Dennis**[2,*]

[1]Department of Biological Sciences, University at Buffalo, 109 Cooke Hall, Buffalo, NY 14260, USA

[2]Department of Biochemistry & Molecular Medicine, Genome Center, and MIND Institute, University of California, Davis, CA 95616, USA

Abstract

Advances in sequencing technologies have enabled the comparison of high-quality genomes of diverse primate species, revealing vast amounts of divergence due to structural variation. Given their large size, structural variants (SVs) can simultaneously alter the function and regulation of multiple genes. Studies estimate that collectively more than 3.5% of the genome is divergent in humans versus other great apes, impacting thousands of genes. Functional genomics and gene-editing tools in various model systems recently emerged as an exciting frontier — investigating the wide-ranging impacts of SVs on molecular, cellular, and systems-level phenotypes. This review examines existing research and identifies future directions to broaden our understanding of the functional roles of SVs on phenotypic innovations and diversity impacting uniquely human features, ranging from cognition to metabolic adaptations.

Introduction

In the millions of years since modern humans diverged from a common ancestor with chimpanzees, subtle changes in our genomes have resulted in unique adaptations impacting wide-ranging musculoskeletal, brain, and immune response traits, as well as changes in diet and metabolism [1]. Recent advances in genome sequencing technologies have documented the massive genome-wide impact of structural changes, collectively called genomic structural variants (SVs), which include small indels (< 50 bp), as well as larger (> 50 bp) genomic alterations in copy number (e.g. deletions and duplications), insertions (e.g. transposition of repeat elements), and inversions [2,3] (Figure 1). Although less frequent than single-nucleotide changes, SVs collectively account for ~6 times more nucleotide differences between any two humans (31 Mbp vs. 5 Mbp) and represent a significant driver of trait diversity across humans today, as described in recent reviews [4,5]. They can also contribute to trait divergence universal across a species, with possible driver

*Corresponding author: Dennis, Megan Y (mydennis@ucdavis.edu), @meganydennis. *Twitter account: @evobioclio.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

variants identified as human unique and fixed versus other primates [6]. Here, we summarize the literature highlighting the emerging functional roles that human-specific SVs play in adaptation and evolution.

Genomic technologies have accelerated the discovery of functional structural variants

Improvements in genome technologies, such as long-read sequencing (PacBio and Oxford Nanopore Technologies) and scaffolding tools (Hi-C and optical mapping), have allowed the production of high-quality genomes highlighting fixed and polymorphic SVs within and between primate species [6,7] (Figure 1). As a result, the field has relatively accurate estimates of divergence among primates: for example, 2971 deletions and 7649 insertions specific to the human lineage [6], along with 75 fixed human-specific inversions [8]. Nevertheless, there is still uncertainty in divergence across more complex regions, such as segmental duplications (defined as > 1 kbp regions with > 90% identity with another locus in a reference genome [9]) and satellite repeats comprising centromeres. Even though divergence estimates between species involving complex regions are underway, assembly errors and difficulties discerning orthologs/paralogs across high copy and rearranged duplicated regions remain as obstacles. For example, the newest human telomere-to-telomere (T2T) genome has resulted in the discovery of 1621 genes (~32 Mbp of segmental duplication regions) lacking synteny with the chimpanzee reference genome (panTro6 build) [10,11] comprising ~300 multigene families (unpublished). Therefore, we anticipate that the ongoing T2T assembly efforts will continue to provide refinement of divergence estimates among primates, including the segmental duplication and centromeric regions. This is an exciting prospect because, unlike single-nucleotide variants, a single SV event can impact large sections of the genome and immediately alter the function of existing genes, create new genes, alter regulatory sequences, and influence chromatin organization, leading to substantial changes in biological function (Figure 2).

Functional genomic datasets from diverse primate species' primary tissue, cell lines, and derived organoids have enabled the identification of SV-associated divergence of *cis*-regulatory elements [12], altered expression of genes [13,14], and novel transcribed genes [15,16]. For example, an assessment of open chromatin between human and rhesus macaque motor cortex cells suggests that nearly 80% of human-specific candidate *cis*-regulatory elements comprise transposable repeat elements [17]. Variability of transposable elements present in modern humans continues to impact *cis*-regulation, particularly in immune cells [18]. Further, comparisons of chromatin conformation (Hi-C) between human and nonhuman primate lymphoblastoid and induced pluripotent cell lines have shed light on the effects of SVs on genome organization [19–21], such as changes in enhancer–promoter interactions and topologically associating domains (TADs: i.e. self-interacting chromosomal regions). SVs rarely overlap TAD boundaries [22,23], likely due to selective constraints. Nevertheless, those that do can result in TAD disruptions associated with gene expression differences between humans and other hominoids. For example, human-specific inversions with evidence of TAD swapping and altered chromatin looping likely underlie differentially

expressed human and rhesus macaque fetal brain genes [8], with some polymorphic inversions associated with variability in human brain morphology [24].

Human-specific deletions, which fall almost exclusively in noncoding regions and are enriched near genes involved in neural function, have long been proposed as a regulatory driver of the evolution of unique human brain features [25]. Recent studies have used innovative functional genomic approaches to collectively test the impacts of ~16 000 human-specific deletions on enhancer activity and cell phenotypes [26,27] (Figure 3a,b). Using a massively parallel reporter assay (MPRA) to compare small (≤31 bp), cross-species conserved human-specific deletions (hCONDELs) with intact chimpanzee sequences, Xue and colleagues [27] found ~800 sequences altering regulatory activity, including some driving expression of notable gene candidates — *HDAC5*, *LOXL2*, and *PPP2CA* — influencing brain development. Initial studies defining hCONDELs focused on larger (≥50 bp) human-specific deletions [25], while more recent work has considered deletions of sequence not necessarily conserved across species (hDELs). Using CRISPR (clustered regularly interspaced short palindromic repeats) interference (CRISPRi, dCas9 fused to a repressor KRAB domain) to directly test enhancer silencing on pluripotent stem cell proliferation, Fair and colleagues [26] narrowed in on a dozen hDELs impacting human–chimpanzee gene expression divergence. Overall, SV regulatory changes have substantially contributed to human evolution and provide an exciting frontier for functional follow-up studies.

Models of functional structural variants in human evolution

Accumulating evidence shows that SVs impact wide-ranging systems. The SVs found to affect early developmental and neurological traits have tended to be older mutations often fixed in humans [28]. These events arguably are at the core of our speciation event, separating the human lineage from other great apes [29]. In contrast, SVs that shape immune and metabolic traits are more recent, likely reflecting the radiation of our ancestors across the globe [4,6,30,31]. The vast variation in the habitats of ancient and extant humans has likely contributed to the evolution and retained variation of hundreds of metabolic- and immune-system-related SVs [32]. These functional human-specific SVs are biomedically relevant for two broad reasons. Fixed human-specific SVs, often gene duplications and gene family expansions affecting neurodevelopment, led to genomic instability, predisposing these regions to rare diseases linked to *de novo* mutations [33]. In contrast, human-specific SVs that influence metabolic and immune system traits show remarkable variation among individuals and populations, likely shaped by spatially and temporally fluctuating adaptive forces [34]. As a result, these polymorphic SVs often predispose to immune-mediated diseases, such as psoriasis and Crohn’s disease, as well as complex metabolic diseases, such as obesity and diabetes [35]. Thus, elucidating the specific molecular mechanisms through which human-specific SVs affect function provides a crucial framework for investigating evolutionary reasons for human disease in addition to developing clinical tools.

Experimental efforts using model systems can link associations between SVs and molecular function (e.g. gene expression changes) to organismal-level biological processes. Increasingly, researchers have used ‘humanized’ model animals, such as mice, to unearth

the fascinating complexity of mechanisms through which SVs affect traits (Figure 3a). One recent interesting example assessed diverse transcripts across apes, discovering an intronic *Alu* insertion impacting splicing of *TBXT*, a gene previously implicated in tail formation. Transgenic studies in mice connected the variant with altered gene function, suggesting a potential role in tail loss among primates [36].

Brain evolution

In addition to hDELs, considerable work has explored the role that gene duplications contribute to unique *Homo* brain features. Collectively, human-specific gene duplications are enriched for neurological functions and reside at genomic hotspots susceptible to nonallelic homologous recombination associated with neurocognitive conditions [37,38]. New duplicated paralogs can increase gene dosage or antagonistically interact with ancestral paralogs, impacting conserved functions or resulting in novel functions with altered expression patterns [39]. Over the past 10 years, functional studies in cerebral organoids and diverse animal models, ranging from mice to monkeys, have highlighted the putative roles of human-specific genes in neocortex development, including functions in synaptogenesis and neuronal proliferation (detailed in previous reviews [5,40]). Two of the most well-studied examples, Rho-GTPase activating protein 11B (*ARHGAP11B*) [41] and Slit-Robo activating protein 2C (*SRGAP2C*) [37,42], result in marked improvements in memory and learning in adult ‘humanized’ transgenic mice likely as a result of increased neocortical sizes [43] and cortical connectivity and circuit function changes [44], respectively. While much work has focused on characterizing novel, human-specific paralogs with apparent immediate phenotypic effects, shared primate paralogs that accrue sporadic mutations in the hominin lineage can also contribute to evolutionarily new features.

Notch signaling—Segmental duplications are subject to elevated mutation rates and interlocus gene conversion [45]. As a consequence, evolutionary innovations through functional changes to existing, older members of multigene families are common. Partial duplications (*NOTCH2NL* paralogs) of the highly conserved Notch receptor 2 (*NOTCH2*) signaling gene represent interesting examples of this phenomenon. *NOTCH2* is essential in maintaining the progenitor pool of radial glia cells [46]. While *NOTCH2NL* exists as nonfunctional pseudogenes in chimpanzees and gorillas, a single shared paralog was likely ‘revived’ along the human lineage via interlocus gene conversion with the ancestral *NOTCH2*. This was followed by two additional duplications producing three protein-producing truncated paralogs on chromosome 1q21.1 [47]. Ectopic expression of *NOTCH2NL* activates the NOTCH signaling pathway, resulting in delayed neuronal differentiation and increased proliferation in mouse cortical spheroids, human cortical progenitors, and the developing mouse brain [48–50]. As a result, the unusually complex gain-of-function event and subsequent duplication of human-specific *NOTCH2Ls* may partially underlie neocortex expansion of the human brain.

Synaptic function—Gene duplications can diverge in expression over evolutionary time, gaining novel organismal-level functions in a species-specific manner. For example, the Leucine-Rich Repeat Containing 37 (*LRRC37*) gene family has evolved into many paralogs shared within simian and hominid species [51]. Recent work has shown that hominid-

specific *LRRC37B* localizes uniquely to the axon initiation segment of human cortical pyramidal neurons (but not in chimpanzees), leading to reduced neuronal excitability through interactions with sodium ion channels when ectopically expressed in mice [52]. These results may explain the distinct electrophysiological properties observed at the axon initiation segment in humans versus rodents [53]. Unlike other paralogs in this gene family, both *LRRC37B* and its human-specific paralog, *LRRC37A3*, are nearly fixed in modern humans and exhibit high neuronal expression patterns in single-cell transcriptomic data. Collectively, these results strongly suggest that both genes may have adaptively evolved in the human lineage and affect brain function.

Metabolic adaptations

The human diet has changed dramatically over our species' evolution. Feast and famine cycles and migrations into new environments with varying resources define our past. Thus, there is a tremendous dietary range among extant and past human populations, from the fat-rich sustenance of Inuit to starch-dominated cuisines of agricultural societies. Recent studies have highlighted the mechanisms through which functional effects of SVs play a central role in human adaptations to diverse diets and oscillations in resource availability.

Amylase—One of the most well-known examples of potential dietary adaptation is the human-specific salivary amylase (*AMY1*) gene duplication. *AMY1* encodes for the amylase enzyme responsible for starch digestion. Chimpanzees have only one haploid copy of this gene, whereas extant humans have one to nine haploid copies based on recent assemblies [54,55], with higher copies associated with agricultural diets [56]. The amylase locus exemplifies the challenges in studying the exact mechanism through which SVs affect metabolic function. Previous studies have connected *AMY1* copy number diversity with obesity [57] and gastrointestinal microbiome composition [58], though some of these associations are disputed [59] and almost certainly context-dependent [60]. Further, the mutational landscape of the amylase locus is complex. Overlapping segmental duplications and multiple retrotransposons underlie recurrent nonallelic homologous recombination and microhomology-mediated break-induced replication events, generating inversions, deletions, and duplications in addition to gene copy number variations with unknown functional consequences. Complicating the direct use of the mouse as a model to investigate the human amylase locus, the salivary amylase in the mouse lineage has convergently evolved through lineage-specific duplications [61]. The exact functional role of salivary amylase gene copy number in recent human evolution remains one of the most interesting mysteries in the field.

Growth hormone receptor—Another recent study described the evolutionary impact of the polymorphic deletion of the third exon in the growth hormone receptor gene (*GHR*) in the human lineage. The deletion (*GHRd3*) exhibits varied allele frequencies ranging from 5% to 25% among human populations and generates a shorter *GHR* isoform (Figure 2 iii and Figure 3b), which is associated with several human phenotypic traits, including altered birth weight, puberty onset, lifespan, and metabolic activity [62]. A population genetic analysis showed near fixation of this deletion in early human evolution, followed by a rapid, adaptive decline in the last 30 000 years. Evidence from recent mouse models reveals sex- and environment-dependent effects of *GHRd3*, leading to female-like transcriptomic, lipidomic,

and growth phenotypes in male mice under caloric restriction. Further, analysis of the downstream effects of *GHRd3* in the context of *BCL6* gene function in mice shows potential loss of immune response in males to certain bacterial infections. Last but not least, *GHRd3* may be protective against kwashiorkor (i.e. bilateral extremity swelling due to severe protein malnutrition). Combined, the population genetic analysis, trait associations, and functional insights from mouse experiments paint a complex evolutionary picture, where *GHRd3* has evolved in a trade-off between starvation resistance and defense against pathogens. As a result, this hDEL has oscillated in frequency during human evolution due to fluctuating selective pressures over time, likely in response to nutritional stress/malnutrition [63].

Immune response

Common functional human variation markedly overlaps with immune-related genomic regions. Some of these immune-related SVs are ancient and have remained polymorphic since human–Neanderthal and even human–chimpanzee split due to balancing selection to counter the pressures from fast-evolving pathogens [64]. Within this context, an emerging hypothesis is that rapidly evolving structural variation affecting immune-related regions, including *HLA*, *LCE3*, and immunoglobulin gene families, have been evolving under frequency-based and diversifying selection.

Mucins—Mucin genes are categorized based on their function (i.e. coding O-glycosylated proteins) rather than a common evolutionary ancestor. They all harbor exonic tandem repeats, which are enriched for codons corresponding to proline, tryptophan, and serine amino acids, which underlie the glycosylation potential of the mucin proteins. Through the attached O-glycans, mucins often interact with commensal and pathogenic microbes, in epithelial surfaces and bodily fluids, including saliva. As such, they are an integral part of the immune system. Long-read sequencing-based variation maps have revealed a surprising level of copy-number variation of the exonic mucin repeats. For example, more than ~5% of genic novel sequences in African pangenomes affect mucin genes [65]. Another study identified 15 instances of evolutionary convergence, where novel mucins recurrently evolve from proline-rich proteins by gaining densely O-glycosylated exonic repeat domains and remain copy-number variable among mammals, affecting glycosylation potential in different tissues [66]. In parallel, variation in mucin copy-number variation is strongly associated with inflammatory diseases [67,68] and microbiome composition [69,70]. Overall, human-specific mucin repeat variation and its impact on glycome in different tissues is an excellent area of research concerning the effect of rapidly evolving pathogens unique to our species.

Future directions

The unprecedented resolution of the genomic and functional impact of SVs in humans and nonhuman primates has allowed us to construct complex evolutionary models of human evolution (Figure 4). As T2T assemblies and pangenomes of diverse primates and humans become routine [11,71,72], improved discovery of variation at recalcitrant regions — including satellite repeats comprising centromeres and acrocentric regions — will allow us to explore the most quickly evolving parts of our genomes and connect them with human traits and diseases. Increasing the number of genomes across species will also delineate

variants that are fixed and divergent between primate species that might contribute to human universal features (e.g. cognitive abilities) from polymorphic within species that can impact diverse phenotypes responsive to varied environmental factors (e.g. metabolism and immunity). As we sequence more individuals, divergence estimates will decrease due to better estimates of fixed versus polymorphic versus incomplete lineage sorting, while diversity estimates will increase due to the identification of rare variants (Figure 1). Better connecting variants with molecular effects is becoming possible with long-read epigenetic and transcriptomic datasets, which can now accurately parse paralog expression differences of nearly identical genes [73,74]. In this review, we have highlighted how functional studies that model phenotypes have successfully used mammalian models, albeit on a relatively small scale. These studies often model whole gene deletions or duplications by overexpressing or knocking out genes. Even though these approaches can be extremely informative in certain cases, a majority of SVs likely affect function in more subtle ways, as exemplified in Figure 2. Therefore, an exciting future direction for understanding the subtler functional impacts of SVs, such as effects on splicing or enhancer activity, involves introducing precise SV breakpoints in model organisms through gene editing. Considering the large numbers of genes putatively impacted by SVs, modeling their functions at scale using organoids presents an exciting opportunity, especially considering expected improvements in mutational editing efficiencies and reduced variability across replicates. The use of higher-throughput nonmammalian organisms, such as zebrafish, also offers a compelling avenue to test interactions between gene duplicate paralogs (e.g. antagonism or neofunctionalization) and understand the functions of uncharacterized ancestral paralogs. The era of SV exploration is here, and the next major frontier is elucidating their functions at molecular, cellular, and organismal levels. We are excited to discover the hidden clues about human evolution that are surely waiting to be uncovered in the complex depths of our genome.

Acknowledgements

Because of space constraints of this mini-review, we were not able to highlight all important studies. We thank Tonia Brown for substantive edits to the manuscript. This work was supported, in part, by grants from the U.S. National Science Foundation, USA (2049947 and 2123284 to O.G. as well as a CAREER Award 2145885 to M.Y.D.) and the U.S. National Institutes of Health, National Institute of Mental Health, USA (R01MH132818 to M.Y.D.). All figures were created with BioRender.com.

Data Availability

No data were used for the research described in the article.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest

•• of outstanding interest

1•. Pollen AA, Kilik U, Lowe CB, Camp JG: Human-specific genetics: new tools to explore the molecular and cellular basis of human evolution. Nat Rev Genet 2023, 24:687–711. [PubMed:

36737647] This paper presents a comprehensive review of known genetic factors driving human-specific traits and the cellular and molecular tools necessary to characterize their functions.

2•. Porubsky D, Eichler EE: A 25-year odyssey of genomic technology advances and structural variant discovery. Cell 2024, 187:1024–1037. [PubMed: 38290514] This review outlines the technology advances over the last two decades enabling SV discovery and their implications to scientific understanding.

3. Ho SS, Urban AE, Mills RE: Structural variation in the sequencing era. Nat Rev Genet 2020, 21:171–189. [PubMed: 31729472]

4•. Hollox EJ, Zuccherato LW, Tucci S: Genome structural variation in human evolution. Trends Genet 2022, 38:45–58. [PubMed: 34284881] A review of SVs and human evolution.

5•. Soto DC, Uribe-Salazar JM, Shew CJ, Sekar A, McGinty SP, Dennis MY: Genomic structural variation: a complex but important driver of human evolution. Am J Biol Anthr 2023, 181:118–144.A review of SV roles in human evolution with an emphasis on human-specific duplications and brain evolution.

6••. Mao Y, Harvey WT, Porubsky D, Munson KM, Hoekzema K, Lewis AP, Audano PA, Rozanski A, Yang X, Zhang S, et al. : Structurally divergent and recurrently mutated regions of primate genomes. Cell 2024, 187:1547–1562.e13. [PubMed: 38428424] This paper provides a comprehensive compendium of lineage-specific SVs using complete genome assemblies of eight diverse primate species complemented with long- and short-read datasets from additional individuals to delineate polymorphic from fixed variants. To date, it represents the most complete estimate of human-specific fixed deletions and insertions.

7. Kronenberg ZN, Fiddes IT, Gordon D, Murali S, Cantsilieris S, Meyerson OS, Underwood JG, Nelson BJ, Chaisson MJP, Dougherty ML, et al. : High-resolution comparative analysis of great ape genomes. Science 2018, 360:eaar6343.

8. Ding W, Li X, Zhang J, Ji M, Zhang M, Zhong X, Cao Y, Liu X, Li C, Xiao C, et al. : Adaptive functions of structural variants in human brain development. Sci Adv 2024, 10:eadl4600.

9. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE: Recent segmental duplications in the human genome. Science 2002, 297:1003–1007. [PubMed: 12169732]

10••. Vollger MR, Guitart X, Dishuck PC, Mercuri L, Harvey WT, Gershman A, Diekhans M, Sulovari A, Munson KM, Lewis AP, et al. : Segmental duplications and their variation in a complete human genome. Science 2022, 376:eabj6965.An assessment of the segmental duplication landscape in a complete human genome assembly, identifying human-specific genes.

11••. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, et al. : The complete sequence of a human genome. Science 2022, 376:44–53. [PubMed: 35357919] This paper describes the sequence and assembly of the first T2T complete human genome (T2T-CHM13), derived from a homozygous hydatidiform mole cell line.

12. García-Pérez R, Esteller-Cucala P, Mas G, Lobón I, Di Carlo V, Riera M, Kuhlwilm M, Navarro A, Blancher A, Di Croce L, et al. : Epigenomic profiling of primate lymphoblastoid cell lines reveals the evolutionary patterns of epigenetic activities in gene regulatory architectures. Nat Commun 2021, 12:3116. [PubMed: 34035253]

13. Muchnik SK, Lorente-Galdos B, Santpere G, Sestan N: Modeling the evolution of human brain development using organoids. Cell 2019, 179:1250–1253. [PubMed: 31778651]

14. Pollen AA, Bhaduri A, Andrews MG, Nowakowski TJ, Meyerson OS, Mostajo-Radji MA, Di Lullo E, Alvarado B, Bedolli M, Dougherty ML, et al. : Establishing cerebral organoids as models of human-specific brain evolution. Cell 2019, 176:743–756.e17. [PubMed: 30735633]

15. Ferrández-Peral L, Zhan X, Alvarez-Estape M, Chiva C, Esteller-Cucala P, García-Pérez R, Julià E, Lizano E, Fornas Ò, Sabidó E, et al. : Transcriptome innovations in primates revealed by single-molecule long-read sequencing. Genome Res 2022, 32:1448–1462. [PubMed: 35840341]

16. Dougherty ML, Underwood JG, Nelson BJ, Tseng E, Munson KM, Penn O, Nowakowski TJ, Pollen AA, Eichler EE: Transcriptional fates of human-specific segmental duplications in brain. Genome Res 2018, 28:1566–1576. [PubMed: 30228200]

17•. Zemke NR, Armand EJ, Wang W, Lee S, Zhou J, Li YE, Liu H, Tian W, Nery JR, Castanon RG, et al. : Conserved and divergent gene regulatory programs of the mammalian neocortex. Nature 2023, 624:390–402. [PubMed: 38092918] Comparative analysis of epigenomic and transcriptional data from human and rhesus macaque brain samples. It represents a comprehensive genomic resource to map gene regulatory elements and highlights transposable elements as a driver of gene expression divergence in the developing brain.

18. Chen X, Pacis A, Aracena KA, Gona S, Kwan T, Groza C, Lin YL, Sindeaux R, Yotova V, Pramatarova A, et al. : Transposable elements are associated with the variable response to influenza infection. Cell Genom 2023, 3:100292. [PubMed: 37228757]

19. Soto DC, Shew C, Mastoras M, Schmidt JM, Sahasrabudhe R, Kaya G, Andrés AM, Dennis MY: Identification of structural variation in chimpanzees using optical mapping and nanopore sequencing. Genes 2020, 11:276. [PubMed: 32143403]

20. Eres IE, Luo K, Hsiao CJ, Blake LE, Gilad Y: Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. PLoS Genet 2019, 15:e1008278. [PubMed: 31323043]

21. Maggiolini FAM, Sanders AD, Shew CJ, Sulovari A, Mao Y, Puig M, Catacchio CR, Dellino M, Palmisano D, Mercuri L, et al. : Single-cell strand sequencing of a macaque genome reveals multiple nested inversions and breakpoint reuse during primate evolution. Genome Res 2020, 30:1680–1693. [PubMed: 33093070]

22. Porubsky D, Sanders AD, Höps W, Hsieh P, Sulovari A, Li R, Mercuri L, Sorensen M, Murali SC, Gordon D, et al. : Recurrent inversion toggling and great ape genome evolution. Nat Genet 2020, 52:849–858. [PubMed: 32541924]

23. Fudenberg G, Pollard KS: Chromatin features constrain structural variation across evolutionary timescales. Proc Natl Acad Sci USA 2019, 116:2175–2180. [PubMed: 30659153]

24. Wang H, Makowski C, Zhang Y, Qi A, Kaufmann T, Smeland OB, Fiecas M, Yang J, Visscher PM, Chen C-H: Chromosomal inversion polymorphisms shape human brain morphology. Cell Rep 2023, 42:112896. [PubMed: 37505983]

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

26. Fair T, Pavlovic BJ, Schaefer NK, Pollen AA: Mapping cis-and trans-regulatory target genes of human-specific deletions. bioRxiv 2023, 10.1101/2023.12.27.573461

27. Xue JR, Mackay-Smith A, Mouri K, Garcia MF, Dong MX, Akers JF, Noble M, Li X, Zoonomia Consortium, Lindblad-Toh K, et al. : The functional and evolutionary impacts of human-specific deletions in conserved elements. Science 2023, 380:eabn2253.

28. Shao Y, Zhou L, Li F, Zhao L, Zhang B-L, Shao F, Chen J-W, Chen C-Y, Bi X, Zhuang X-L, et al. : Phylogenomic analyses provide insights into primate evolution. Science 2023, 380:913–924. [PubMed: 37262173]

29. Vanderhaeghen P, Polleux F: Developmental mechanisms underlying the evolution of human cortical circuits. Nat Rev Neurosci 2023, 24:213–232. [PubMed: 36792753]

30. Saitou M, Gokcumen O: An evolutionary perspective on the impact of genomic copy number variation on human health. J Mol Evol 2020, 88:104–119. [PubMed: 31522275]

31. Leffler EM, Band G, Busby GBJ, Kivinen K, Le QS, Clarke GM, Bojang KA, Conway DJ, Jallow M, Sisay-Joof F, et al. : Resistance to malaria through structural variation of red blood cell invasion receptors. Science 2017, 356:eaam6393.

32. Benton ML, Abraham A, LaBella AL, Abbot P, Rokas A, Capra JA: The influence of evolutionary history on human health and disease. Nat Rev Genet 2021, 22:269–283. [PubMed: 33408383]

33. Guitart X, Porubsky D, Yoo D, Dougherty ML, Dishuck PC, Munson KM, Lewis AP, Hoekzema K, Knuth J, Chang S, et al. : Independent expansion, selection and hypervariability of the *TBC1D3* gene family in humans. bioRxiv 2024, 10.1101/2024.03.12.584650

34. Zhang J-Y, Roberts H, Flores DSC, Cutler AJ, Brown AC, Whalley JP, Mielczarek O, Buck D, Lockstone H, Xella B, et al. : Using *de novo* assembly to identify structural variation of eight complex immune system gene regions. PLoS Comput Biol 2021, 17:e1009254. [PubMed: 34343164]

35. Aqil A, Speidel L, Pavlidis P, Gokcumen O: Balancing selection on genomic deletion polymorphisms in humans. Elife 2023, 12:e79111. [PubMed: 36625544]
36. Xia B, Zhang W, Zhao G, Zhang X, Bai J, Brosh R, Wudzinska A, Huang E, Ashe H, Ellis G, et al. : On the genetic basis of tail-loss evolution in humans and apes. Nature 2024, 626:1042–1048. [PubMed: 38418917]
37. Dennis MY, Harshman L, Nelson BJ, Penn O, Cantsilieris S, Huddleston J, Antonacci F, Penewit K, Denman L, Raja A, et al. : The evolution and population diversity of human-specific segmental duplications. Nat Ecol Evol 2017, 1:69. [PubMed: 28580430]
38. Lin Y-L, Gokcumen O: Fine-scale characterization of genomic structural variation in the human genome reveals adaptive and biomedically relevant hotspots. Genome Biol Evol 2019, 11:1136–1151. [PubMed: 30887040]
39. Shew CJ, Carmona-Mora P, Soto DC, Mastoras M, Roberts E, Rosas J, Jagannathan D, Kaya G, O'Geen H, Dennis MY: Diverse molecular mechanisms contribute to differential expression of human duplicated genes. Mol Biol Evol 2021, 38:3060–3077. [PubMed: 34009325]
40. Dennis MY, Eichler EE: Human adaptation and evolution by segmental duplication. Curr Opin Genet Dev 2016, 41:44–52. [PubMed: 27584858]
41. Florio M, Albert M, Taverna E, Namba T, Brandl H, Lewitus E, Haffner C, Sykes A, Wong FK, Peters J, et al. : Human-specific gene *ARHGAP11B* promotes basal progenitor amplification and neocortex expansion. Science 2015, 347:1465–1470. [PubMed: 25721503]
42. Charrier C, Joshi K, Coutinho-Budd J, Kim JE, Lambert N, de Marchena J, Jin WL, Vanderhaeghen P, Ghosh A, Sassa T, et al. : Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. Cell 2012, 149:923–935. [PubMed: 22559944]
43. Xing L, Kubik-Zahorodna A, Namba T, Pinson A, Florio M, Prochazka J, Sarov M, Sedlacek R, Huttner WB: Expression of human-specific *ARHGAP11B* in mice leads to neocortex expansion and increased memory flexibility. EMBO J 2021, 40:e107093. [PubMed: 33938018]
44. Schmidt ERE, Zhao HT, Park JM, Dipoppa M, Monsalve-Mercado MM, Dahan JB, Rodgers CC, Lejeune A, Hillman EMC, Miller KD, et al. : A human-specific modifier of cortical connectivity and circuit function. Nature 2021, 599:640–644. [PubMed: 34707291]
45•. Vollger MR, Dishuck PC, Harvey WT, DeWitt WS, Guitart X, Goldberg ME, Rozanski AN, Lucas J, Asri M, Human Pangenome Reference Consortium, et al. : Increased mutation and gene conversion within human segmental duplications. Nature 2023, 617:325–334. [PubMed: 37165237] Using sequence assemblies created through the Human Pangenome Reference Consortium, authors discovered increased mutation and gene conversion rates across segmental duplications compared with nonduplicated euchromatic genome regions. This study opens up the possibility that as-yet-discovered variants associated with human disease and phenotypes may exist within these difficult-to-assay loci.
46. Nian F-S, Hou P-S: Evolving roles of notch signaling in cortical development. Front Neurosci 2022, 16:844410. [PubMed: 35422684]
47. Fiddes IT, Lodewijk GA, Mooring M, Bosworth CM, Ewing AD, Mantalas GL, Novak AM, van den Bout A, Bishara A, Rosenkrantz JL, et al. : Human-specific *NOTCH2NL* genes affect notch signaling and cortical neurogenesis. Cell 2018, 173:1356–1369.e22. [PubMed: 29856954]
48. Suzuki IK, Gacquer D, Van Heurck R, Kumar D, Wojno M, Bilheu A, Herpoel A, Lambert N, Cheron J, Polleux F, et al. : Human-specific *NOTCH2NL* genes expand cortical neurogenesis through delta/notch regulation. Cell 2018, 173:1370–1384.e16. [PubMed: 29856955]
49. Florio M, Heide M, Pinson A, Brandl H, Albert M, Winkler S, Wimberger WB, Hiller M: Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. ELife 2018, 7:e32332. [PubMed: 29561261]
50. Fiddes IT, Pollen AA, Davis JM, Sikela JM: Paired involvement of human-specific Olduvai domains and NOTCH2NL genes in human brain evolution. Hum Genet 2019, 138:715–721. [PubMed: 31087184]
51. Giannuzzi G, Siswara P, Malig M, Marques-Bonet T, NISC Comparative Sequencing Program, Mullikin JC, Ventura M, Eichler EE: Evolutionary dynamism of the primate *LRRC37* gene family. Genome Res 2013, 23:46–59. [PubMed: 23064749]

52. Libé-Philippot B, Lejeune A, Wierda K, Louros N, Erkol E, Vlaeminck I, Beckers S, Gaspariunaite V, Bilheu A, Konstantoulea K, et al. : LRRC37B is a human modifier of voltage-gated sodium channels and axon excitability in cortical neurons. Cell 2023, 186:5766–5783.e25. [PubMed: 38134874]

53. Testa-Silva G, Verhoog MB, Linaro D, de Kock CPJ, Baayen JC, Meredith RM, De Zeeuw CI, Giugliano M, Mansvelder HD: High bandwidth synaptic communication and frequency tracking in human neocortex. PLoS Biol 2014, 12:e1002007. [PubMed: 25422947]

54•. Yilmaz F, Karageorgiou C, Kim K, Pajic P, Scheer K, Human Genome Structural Variation Consortium, Beck CR, Torregrossa A-M, Lee C, Gokcumen O: Paleolithic gene duplications primed adaptive evolution of human amylase locus upon agriculture. bioRxiv 2024, 10.1101/2023.11.27.568916.This paper provides a comprehensive look at the remarkable human-specific structural complexity of the amylase locus in an evolutionary context.

55•. Bolognini D, Halgren A, Lou RN, Raveane A, Rocha JL, Guarracino A, Soranzo N, Chin J, Garrison E, Sudmant PH: Global diversity, recurrent evolution, and recent selection on amylase structural haplotypes in humans. bioRxiv 2024, 10.1101/2024.02.07.579378.This paper provides a comprehensive look at the remarkable human-specific structural complexity of the amylase locus in an evolutionary context.

56. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, et al. : Diet and the evolution of human amylase gene copy number variation. Nat Genet 2007, 39:1256–1260. [PubMed: 17828263]

57. Falchi M, El-Sayed Moustafa JS, Takousis P, Pesce F, Bonnefond A, Andersson-Assarsson JC, Sudmant PH, Dorajoo R, Al-Shafai MN, Bottolo L, et al. : Low copy number of the salivary amylase gene predisposes to obesity. Nat Genet 2014, 46:492–497. [PubMed: 24686848]

58. Poole AC, Goodrich JK, Youngblut ND, Luque GG, Ruaud A, Sutter JL, Waters JL, Shi Q, El-Hadidi M, Johnson LM, et al. : Human salivary amylase gene copy number impacts oral and gut microbiomes. Cell Host Microbe 2019, 25:553–564.e7. [PubMed: 30974084]

59. Usher CL, Handsaker RE, Esko T, Tuke MA, Weedon MN, Hastie AR, Cao H, Moon JE, Kashin S, Fuchsberger C, et al. : Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. Nat Genet 2015, 47:921–925. [PubMed: 26098870]

60. Hasegawa T, Kakuta M, Yamaguchi R, Sato N, Mikami T, Murashita K, Nakaji S, Itoh K, Imoto S: Impact of salivary and pancreatic amylase gene copy numbers on diabetes, obesity, and functional profiles of microbiome in Northern Japanese population. Sci Rep 2022, 12:7628. [PubMed: 35538098]

61. Pajic P, Pavlidis P, Dean K, Neznanova L, Romano R-A, Garneau D, Daugherity E, Globig A, Ruhl S, Gokcumen O: Independent amylase gene copy number bursts correlate with dietary preferences in mammals. Elife 2019, 8:e44628. [PubMed: 31084707]

62. Brooks AJ, Waters MJ: The growth hormone receptor: mechanism of activation and clinical implications. Nat Rev Endocrinol 2010, 6:515–525. [PubMed: 20664532]

63•. Saitou M, Resendez S, Pradhan AJ, Wu F, Lie NC, Hall NJ, Zhu Q, Reinholdt L, Satta Y, Speidel L, et al. : Sex-specific phenotypic effects and evolutionary history of an ancient polymorphic deletion of the human growth hormone receptor. Sci Adv 2021, 7:eabi4476.This paper integrates population genomics approaches, association studies, and a transgenic mouse model to resolve the specific mechanisms through which growth hormone receptor deletion polymorphism affects function and its role in human evolution.

64. Leffler EM, Gao Z, Pfeifer S, Ségurel L, Auton A, Venn O, Bowden R, Bontrop R, Wall JD, Sella G, et al. : Multiple instances of ancient balancing selection shared between humans and chimpanzees. Science 2013, 339:1578–1582. [PubMed: 23413192]

65. Sherman RM, Forman J, Antonescu V, Puiu D, Daya M, Rafaels N, Boorgula MP, Chavan S, Vergara C, Ortega VE, et al. : Assembly of a pan-genome from deep sequencing of 910 humans of African descent. Nat Genet 2019, 51:30–35. [PubMed: 30455414]

66. Pajic P, Shen S, Qu J, May AJ, Knox S, Ruhl S, Gokcumen O: A mechanism of gene evolution generating mucin function. Sci Adv 2022, 8:eabm8757.

67. Mukamel RE, Handsaker RE, Sherman MA, Barton AR, Zheng Y, McCarroll SA, Loh P-R: Protein-coding repeat polymorphisms strongly shape diverse human phenotypes. Science 2021, 373:1499–1505. [PubMed: 34554798]

68. Plender EG, Prodanov T, Hsieh P, Nizamis E, Harvey WT, Sulovari A, Munson KM, Kaufman EJ, O'Neal WK, Valdmanis PN, et al. : Structural and genetic diversity in the secreted mucins, *MUC5AC* and *MUC5B*. bioRxiv 2024, 10.1101/2024.03.18.585560

69. Hansson GC: Mucins and the microbiome. Annu Rev Biochem 2020, 89:769–793. [PubMed: 32243763]

70. Xu D, Pavlidis P, Taskent RO, Alachiotis N, Flanagan C, DeGiorgio M, Blekhman R, Ruhl S, Gokcumen O: Archaic hominin introgression in Africa contributes to functional salivary *MUC7* genetic variation. Mol Biol Evol 2017, 34:2704–2715. [PubMed: 28957509]

71. Housman G, Tung J: Next-generation primate genomics: new genome assemblies unlock new questions. Cell 2023, 186:5433–5437. [PubMed: 38065076]

72••. Liao W-W, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, Lu S, Lucas JK, Monlong J, Abel HJ, et al. : A draft human pangenome reference. Nature 2023, 617:312–324. [PubMed: 37165242] This paper describes the data release of 47 phased, diploid assemblies representing diverse human individuals. Comparisons across the assemblies discovered novel gene duplication and structural variation not previously captured.

73. Altemose N, Maslan A, Smith OK, Sundararajan K, Brown RR, Mishra R, Detweiler AM, Neff N, Miga KH, Straight AF, et al. : DiMeLo-seq: a long-read, single-molecule method for mapping protein-DNA interactions genome wide. Nat Methods 2022, 19:711–723. [PubMed: 35396487]

74. Gershman A, Sauria MEG, Guitart X, Vollger MR, Hook PW, Hoyt SJ, Jain M, Shumate A, Razaghi R, Koren S, et al. : Epigenetic patterns in a complete human genome. Science 2022, 376:eabj5089.

75. Jeong H, Dishuck PC, Yoo D, Harvey WT, Munson kM, Lewis AP, Kordosky J, Garcia GH, Human Genome Structural Variation Consortium (HGSVC), Yilmaz F, et al. : Structural polymorphism and diversity of human segmental duplications. bioRxiv 2024, 10.1101/2024.06.04.597452

76. Dennis MY, Nuttle X, Sudmant PH, Antonacci F, Graves TA, Nefedov M, Rosenfeld JA, Sajjadian S, Malig M, Kotkiewicz H, et al. : Evolution of human-specific neural *SRGAP2* genes by incomplete segmental duplications. Cell 2012, 149:912–922. [PubMed: 22559943]

77. Huttner W: Neocortical Neurogenesis in Development and Evolution. John Wiley & Sons; 2023.

78. Schmidt ERE, Zhao HT, Park JM, Dipoppa M, Monsalve-Mercado MM, Dahan JB, Rodgers CC, Lejeune A, Hillman EMC, Miller KD, et al. : A human-specific modifier of cortical connectivity and circuit function. Nature 2021, 599:640–644. [PubMed: 34707291]

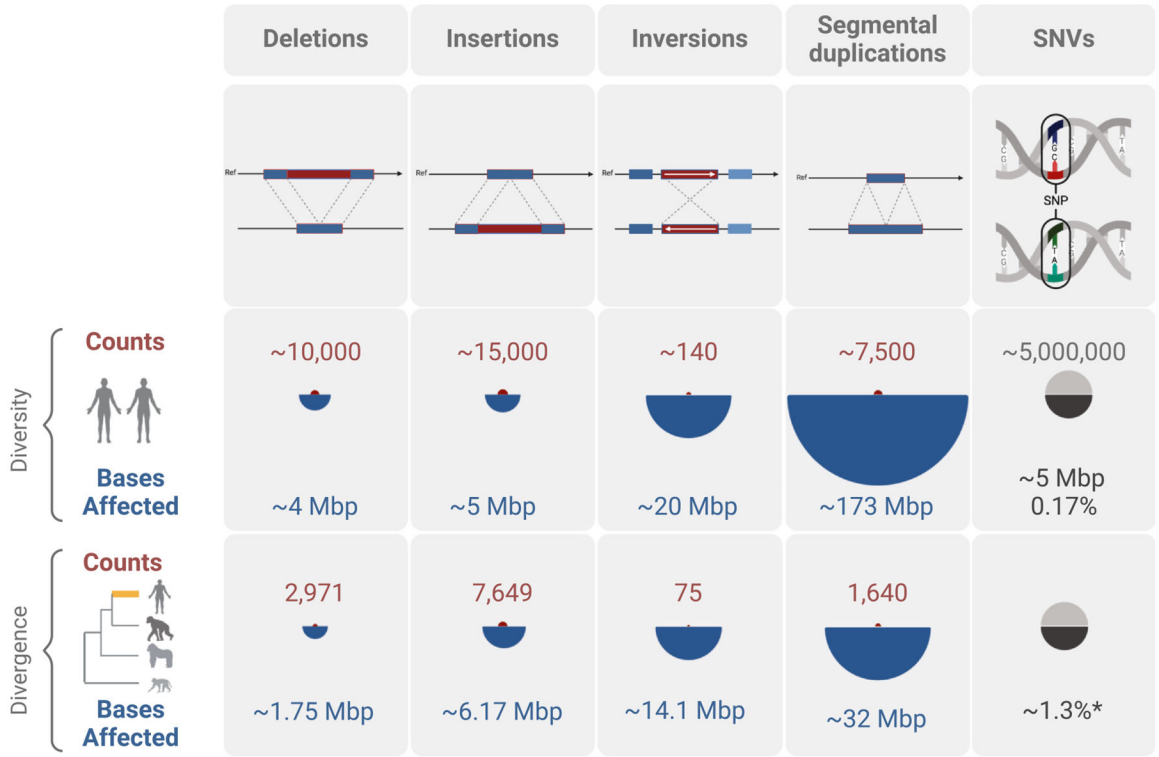

Figure 1.

Genomic landscape of structural variation. The figure depicts two levels of comparison: the upper section (‘Diversity’) captures the average variability between a diploid human genome and a human reference genome. Here, the figure captures both nucleotide and SVs reported to be polymorphic within humans. In contrast, the lower section (‘Divergence’) highlights human-specific genomic variants/fixed SVs in the human lineage compared to other available nonhuman primate genomes. Thus, the genetic variants shown here represent the genetic differences between species or lineages that have accumulated independently since their split from a common ancestor. It is important to note that although divergence primarily refers to fixed changes, the inclusion of additional genomes for a particular species can influence the classification of variants as polymorphic or fixed. For instance, Ding et al. [8] detected 75 human-specific inversions, whereas previously, 130 were thought to be human-specific [7]. The inclusion of new genomes allows us to investigate whether observed SVs are fixed or polymorphic in a given lineage as a result of incomplete lineage sorting, systematically reducing divergent variants as additional individuals are added to analyses. By considering both fixed and polymorphic changes, we gain a comprehensive view of the genetic landscape within populations (diversity) and between species or lineages (divergence). The SVs indicated are deletions, insertions, inversions, and segmental duplications, while single-nucleotide variants (SNVs) are also included in the figure for comparative purposes. Overlaps exist between insertions and segmental duplications, but based on the methodological differences in their identification, we have chosen to include both SV types. SV counts are shown in red, and the affected bases in blue, whereas SNV counts and the bases affected are depicted in grayscale. Regarding the divergent segmental duplications, the count reflects the total number of genic segmental duplications

identified in the human T2T-CHM13 genome compared to the chimpanzee genome [10]. In contrast, the segmental duplications within humans, reflecting diversity, are based on the total number of segmental duplications reported by Jeong et al. [75]. The counts and bases affected correspond to the total segmental duplications detected, not just the genic ones. The SNVs reported as divergent represent the percentage of nucleotide divergence estimated between humans and chimpanzees*, while the divergent SVs reported correspond to the fixed, human-specific deletions, insertions [6], and inversions [8]. It is worth noting that, unlike SNVs, the extent to which different types of SVs affect primate genomes is yet to be fully resolved and numerous SVs exist beyond those listed here.

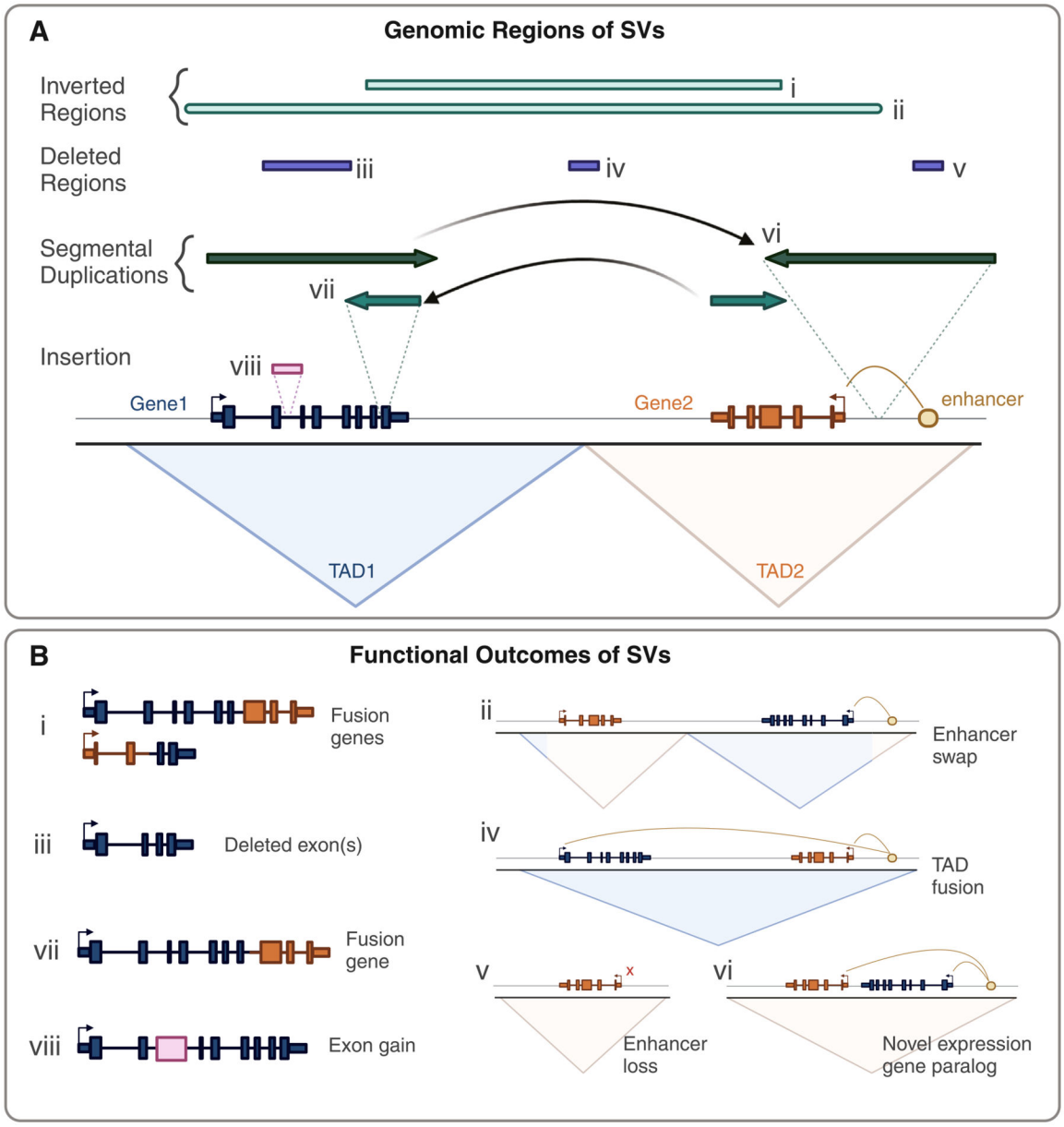

Figure 2.

Possible consequences of SVs impacting gene function and regulation. Numbered SVs (i–viii) in panel **(a)**, which summarizes the genomic regions of SVs, indicate the genomic context with coding (Gene 1 and Gene 2) and regulatory (enhancer regulating Gene 2) sequences and topologically associating domains (TAD1 and TAD2). The lower panel **(b)** shows the putative functional outcomes of each numbered SV on gene function and regulation. Studies have demonstrated the functional impact of SVs on, for example, chromatin organization [8,19–21], splicing [36], gene duplications with novel expression patterns [39,41,52], and exonic deletions [63].

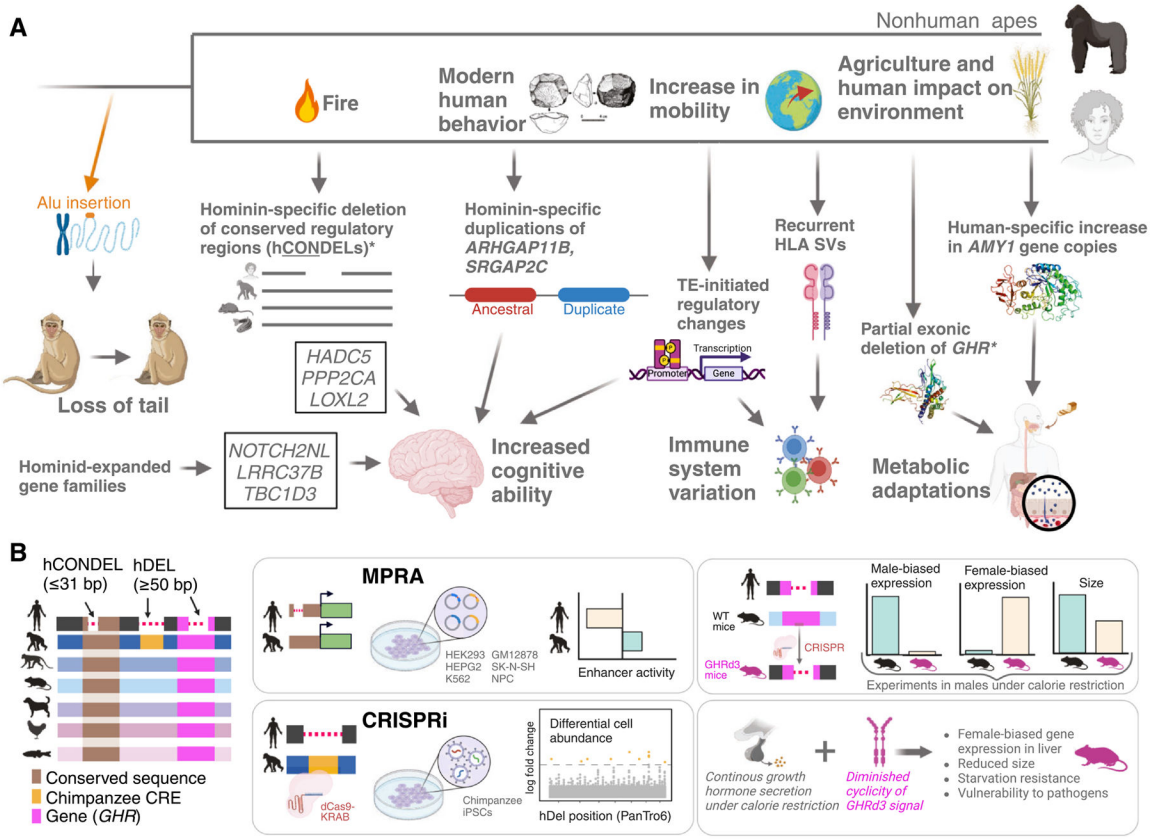

Figure 3.

Recent examples of functional SVs and their roles in evolution. **(a)** We highlight the examples of human-specific SVs that affect crucial human traits and biological function [17,18,33,36,41–43,47,48,52,76–78]. Specific biological processes shaped by SVs differ across evolutionary time, likely because the nature of the adaptive-pressures changes during human evolution. Notably, more recent human-specific changes involve the immune system [66–68,70] and metabolism [54–57,59] and remain variable in human populations. **(b)** Focusing on human-specific deletions (hDELs and hCONDELs), we highlight two studies that characterized noncoding elements and their functions on gene regulation, combining transcriptomic and epigenomic datasets, with functional genomic methods (MPRA [27] and CRISPRi [26]) in diverse human, chimpanzee, and mouse cell lines. A third study [63] tested the functional impacts of a 22-amino hDEL of the gene *GHR* resulting in metabolic phenotypes in a *GHRd3* mouse model. [Glossary: CRE: cis-regulatory element, dCas9-KRAB: inactive form of the Cas9 protein (dCas9) fused to the to a Krüppel-associated box (KRAB), GHR: growth hormone receptor, hCONDELs: human-specific deletions in conserved regions, hDELs: human-specific deletions, HLA: human leukocyte antigens - genes in major histocompatibility complexes (MHC), MPRA: massively parallel reporter assay, TE: transposable elements].

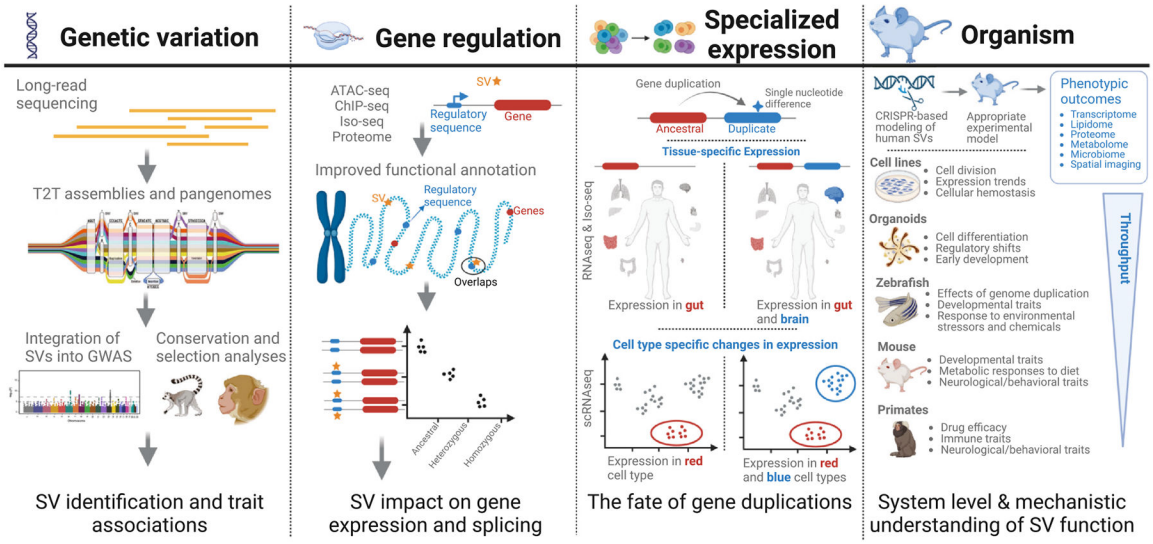

Figure 4.

The investigative frontiers of human-specific SVs and uncovering their functions.