



## OPEN ACCESS

## EDITED BY

Qiong Zhang,  
Albert Einstein College of Medicine,  
United States

## REVIEWED BY

Anna J. Jasinska,  
University of Pittsburgh, United States  
Mingyuan Li,  
Johns Hopkins University, United States  
Yong Shao,  
Kunming Institute of Zoology (CAS),  
China

## \*CORRESPONDENCE

Wen-Hua Wei,  
✉ wenhuaatchristchurch@gmail.com

## SPECIALTY SECTION

This article was submitted to Genomic  
Analysis, a section of the journal  
Frontiers in Bioinformatics

RECEIVED 08 February 2023

ACCEPTED 07 March 2023

PUBLISHED 28 March 2023

## CITATION

Wei W-H and Guo H (2023), Leveraging  
primate-specific genomic information  
for genetic studies of complex diseases.  
*Front. Bioinform.* 3:1161167.  
doi: 10.3389/fbinf.2023.1161167

## COPYRIGHT

© 2023 Wei and Guo. This is an open-  
access article distributed under the terms  
of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original author(s)  
and the copyright owner(s) are credited  
and that the original publication in this  
journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Leveraging primate-specific genomic information for genetic studies of complex diseases

Wen-Hua Wei<sup>1,2,3\*</sup> and Hui Guo<sup>3</sup>

<sup>1</sup>Department of Women's and Children's Health, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand, <sup>2</sup>Department of Biochemistry, University of Otago, Dunedin, New Zealand, <sup>3</sup>Centre for Biostatistics, School of Health Sciences, The University of Manchester, Manchester, United Kingdom

Genomic changes specific to higher primates are regarded as primate-specific genomic information (PSI). Using PSI to inform genetic studies is highly desirable but hampered by three factors: heterogeneity among PSI studies, lack of integrated profiles of the identified PSI elements and dearth of relevant functional information. We report a database of 19,767 PSI elements collated from nine types of brain-related studies, which form 19,473 non-overlapping PSI regions that distribute unevenly but jointly cover only 0.81% of the genome. About 2.5% of the PSI regions colocalized with variants identified in genome-wide association studies, with disease loci more likely colocalized than quantitative trait loci ( $p = 1.6 \times 10^{-5}$ ), particularly in regions without obvious regulatory roles. We further showed an *LRP4* exemplar region with PSI elements orchestrated with common and rare disease variants and other functional elements. Our results render PSI elements as a valuable source to inform genetic studies of complex diseases.

## KEYWORDS

complex disease, colocalization, genetic regulation, pleiotropy, primate-specific genomic information (PSI)

## 1 Introduction

Using human-specific genomic changes to inform studying human-specific features has been a hot topic (Pollen et al., 2023) since the first discovery of human accelerated regions (HARs) that were highly conserved across vertebrates but with an accelerated rate of evolution from chimpanzees to humans in nucleotide substitution (Pollard et al., 2006). The roles of HARs have been confirmed in regulation of human-specific morphological features such as limb and brain development (Prabhakar et al., 2008; Boyd et al., 2015). Additional human-specific genomic changes have been uncovered including human accelerated DNase I hypersensitive sites (haDHS) (Gittelman et al., 2015), human biased cis-regulatory elements (Prescott et al., 2015), Hominin-specific gene regulatory elements (hsGRE) (Castelijns et al., 2020), and large structural changes such as human-specific insertions (hsInsert) that tend to be fixed in human populations (McLean et al., 2011; Kronenberg et al., 2018).

Furthermore, researchers have uncovered substantial genomic changes that are evolutionarily specific to primates (Tay et al., 2009; Vermunt et al., 2016). These primate-specific genomic information (PSI), including the human-specific genomic changes, form an important resource that could potentially shed light on complex mechanisms regulating human phenotypes (Juan et al., 2023). For example, inactivation

of the uricase gene in higher primates has led to a much higher base level of urate concentration than that in species with intact uricase such as mice, and thus may have shaped a special urate-mediated physiology (Lu et al., 2019). Indeed, PSI elements were found in regulatory regions of urate associated genes (Takei et al., 2021) and in the human uricase gene *UOX* (Castelijns et al., 2020).

There are however a few issues limiting PSI applications. First, heterogeneity among PSI elements identified in different studies using evolved definitions and small samples. For example, overlaps exist in only a few of ~2,700 HARs reported from various studies (Levchenko et al., 2018). Second, there is dearth of functional information for the vast majority of the PSI elements, which often requires right tissues or cell types to dissect (Doan et al., 2016; Parikshak et al., 2016; Won et al., 2019) and thus is rather expensive and time consuming. Third, a clear profile of the PSI elements and their interrelations is lacking, e.g., genomic distribution, frequencies in populations.

Another difficulty is to clarify potential roles of PSI elements in regulating human diseases, given that PSI elements are concerned at the species level rather than within-species or individual levels. PSI elements tend to be subject to positive selection leading to fixation in human populations (McLean et al., 2011; Kronenberg et al., 2018) while such positive selection could be reverted to negative selection during human evolution (e.g., adaptation to different environments) (Doan et al., 2016; Levchenko et al., 2018), leading to phenotypic variations including diseases. Furthermore, mutations can happen in PSI elements where deleterious variants may be associated with diseases such as neural disorders underlying the emergence of the human brain (Doan et al., 2016; Castelijns et al., 2020; Takei et al., 2021). These attributes together render PSI elements as favorable targets for new insights into disease associations and primate-specific regulatory mechanisms.

Here we characterize PSI elements collated from a range of publications mostly on brain-related features and their colocalization with common variants identified in genome-wide association studies (GWASs). We quote genomic positions in GRCh37/hg19 throughout.

## 2 Materials and methods

To prepare for a whole genome sequencing based project aimed to define human specific genetic variants in brain developmental disorders, journal articles studying primate-specific and/or human-specific genomic features were reviewed and selected based on relevance and usefulness for annotating sequencing variants. We then extracted data from each selected journal article and converted the genome coordinates to GRCh37/hg19 when necessary.

The PSI elements collated were first sorted by chromosome locations and merged into PSI regions without overlaps with each other (Supplementary Table S1). Subsequently, distribution and summary statistics of PSI regions were generated. The PSI regions were then used as a map to examine colocalization of common variants identified in GWASs of 10 quantitative traits and 15 dichotomous disease traits outlined in Supplementary Table S2. Mapping GWAS variants to PSI regions and statistical tests were performed using R (R Core Team, R, 2018) (<https://www.R-project.org>, v4.1.1). Chi-square tests were used to examine hypothesis that no difference between the

expected and the observed frequencies in one or more categories of a contingency table, or hypothesis of independence between two distributions. For each GWAS, genome-wide significant ( $p < 5 \times 10^{-8}$  or  $\log_{10}$ Bayes Factor  $>6$  as appropriate in the study) variants were extracted and each tested for mapping to the PSI regions and recorded with the PSI type if mapped using a self-developed R script (available on request). Then the numbers of independent loci were derived from the input and the mapped variants respectively by counting linkage disequilibrium blocks ( $r^2 > 0.8$ ) associated with the lead associated variants that were mutually independent and reported in the original paper. For example, if five mapped variants are within a linkage disequilibrium block of a GWAS lead associated variant, they will be counted as one mapped locus. Mapping rate was then calculated as percentage of the number of mapped loci out of the number of GWAS loci. The GTEx Portal (Lonsdale et al., 2013) (<https://gtexportal.org/home/>) was used to retrieve functional information such as expression quantitative trait loci (eQTL) or splicing quantitative trait loci (sQTL) for variants of interest.

## 3 Results

We studied 19,767 PSI elements identified from nine studies each with a different PSI type based on DNA sequences and/or brain related tissues or cell types (Table 1). These PSI elements formed 19,473 non-overlapping PSI regions that are generally small (Supplementary Table S1), with an average length of 1,268 base pairs, together cover only 0.81% of the genome. Only 226 PSI regions (1.2%) were derived from multiple PSI elements confirming fairly limited overlaps between PSI types. The PSI regions distributed unevenly by chromosome ( $p < 2.2 \times 10^{-16}$ , Chi-squared test accounting for chromosomal lengths), where the chromosomal length coverage by PSI regions fluctuated around 0.8% but was above 1% in chromosomes 11 and 20 and as low as 0.5% in chromosomes 21, 22, and X (Figure 1A).

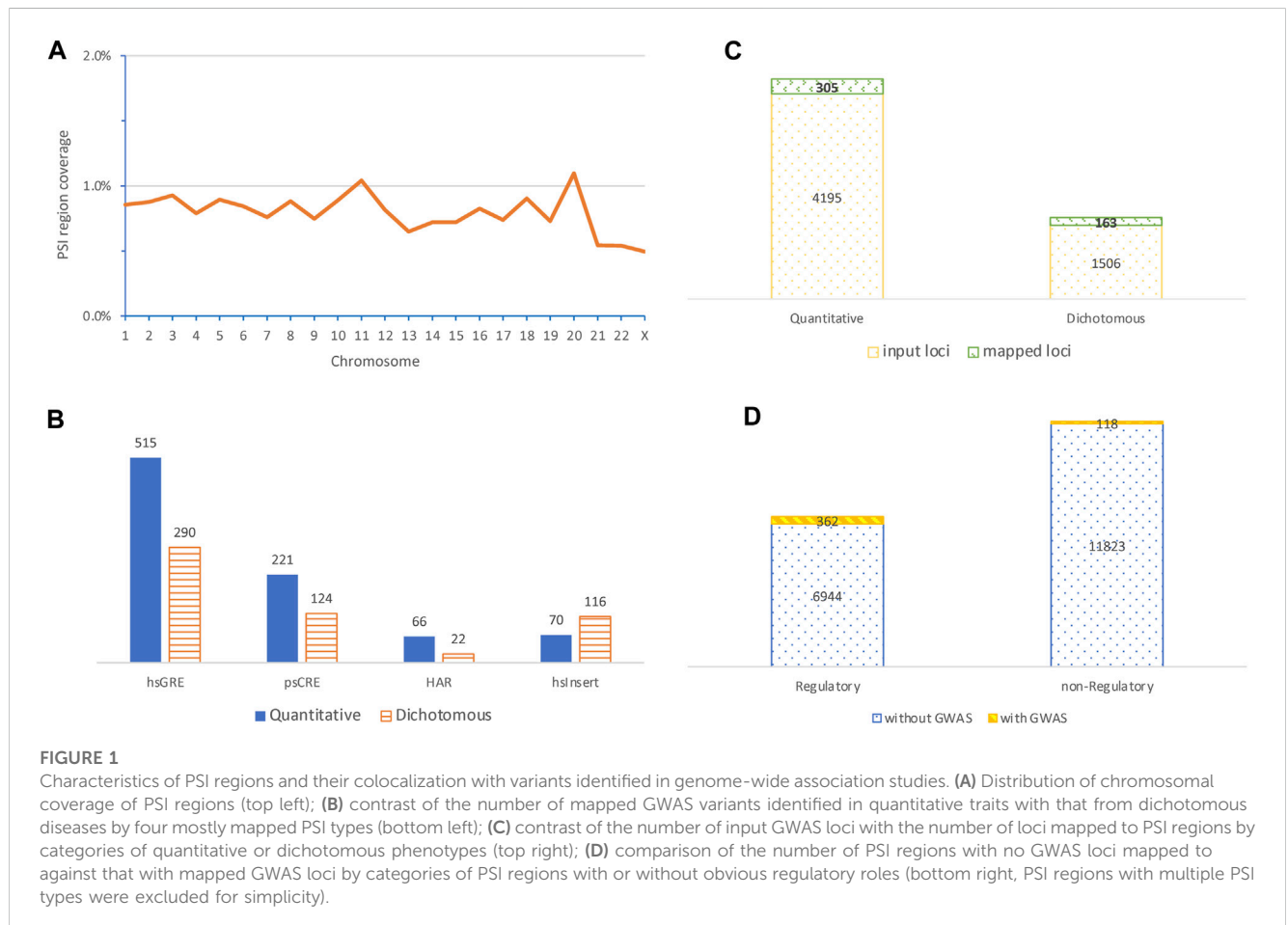
Of the 19,473 PSI regions, 493 (~2.5%) were mapped by 1,499 (representing 468 independent loci) out of 253,200 genome-wide significant variants (representing 5,701 independent loci) identified in GWASs of 10 quantitative and 15 dichotomous disease traits, giving a mapping rate of 0.6% (or 8.2%) at the variant (or locus) level (Supplementary Table S2). Of the 1,499 mapped variants, 332 (~22%) were pleiotropic and associated with multiple traits (Supplementary Table S3). At the variant level, the top mapped PSI types were hsGRE, psCRE, hsInsert, and HAR (Figure 1B), where hsInsert was the only one mapped with more disease variants than quantitative trait variants and was significantly different from the remaining ( $p = 9.3 \times 10^{-12}$ , Chi-squared test). At the locus level, disease loci were more likely mapped to PSI regions than quantitative trait loci ( $p = 1.6 \times 10^{-05}$ , Chi-squared test) (Figure 1C). Considering hsInsert and psTU as the only PSI types without obvious regulatory evidence, a further locus level comparison suggested that the mapped GWAS loci were highly biased towards regulatory regions ( $p = 1.4 \times 10^{-66}$ , Chi-squared test) (Figure 1D). Nevertheless, non-regulatory PSI regions colocalized with 23.9% (118 out of the 493) of the mapped GWAS loci and thus played an important part as well.

Figure 2 showed an exemplar region within the *LRP4* gene harboring GWAS variants and rare pathogenic variants as well as a

TABLE 1 Summary information of the PSI elements<sup>a</sup>.

PSI_type	PSI_description	Brief	Number	Length% (%)	References
1	human accelerated region	HAR	2737	2.8	Doan et al. (2016)
2	human gained enhancers during corticogenesis	HGE	63	0.1	Reilly et al. (2015)
3	human-biased cis-regulatory elements in cranial neural crest cells	hbCRE	1000	0.8	Prescott et al. (2015)
4	primate-specific cis-regulatory elements in brain	psCRE	1488	19.9	Vermunt et al. (2016)
5	recurrent <i>de novo</i> mutations in regulatory elements of neurodevelopment	DNM	31	0.1	Short et al. (2018)
6	human-specific insertion	hsInsert	11874	41.7	Kronenberg et al. (2018)
7	primate-specific transcriptional unit	psTU	131	1.8	Tay et al. (2009)
8	human accelerated DNase I hypersensitive site	haDHS	524	0.7	Gittelman et al. (2015)
9	Hominin-specific gene regulatory element	hsGRE	1919	32.6	Castelijns et al. (2020)

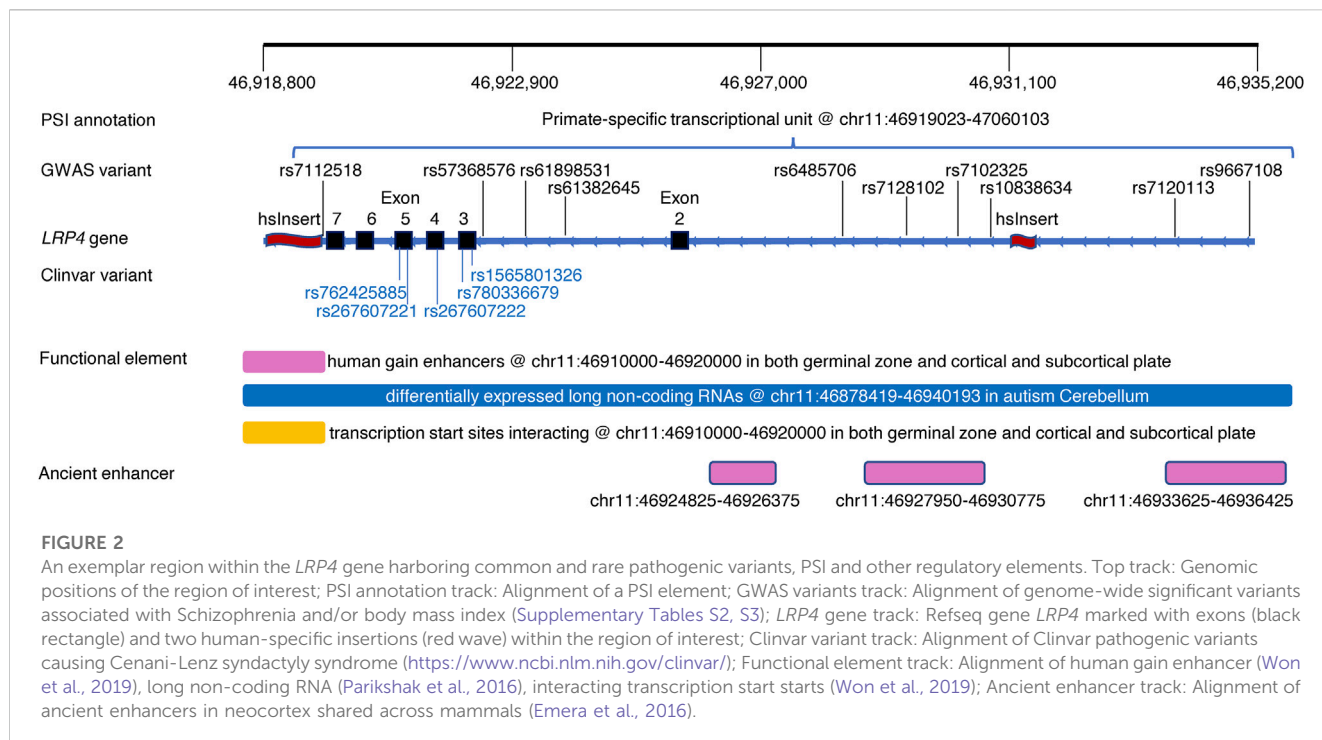
<sup>a</sup>Length%: percentage of the total length per PSI, type over the total length of the PSI, regions (24,708,805 base pairs); brief: Brief name of PSI\_type; ref: Reference of the original study.



psTU overlapped with a hsInsert downstream and another hsInsert upstream. The region also harbors brain related functional elements including human gain enhancers (Reilly et al., 2015), differentially expressed long non-coding RNA (Parikshak et al., 2016), interacting transcription start sites (Won et al., 2016), and three ancient enhancers in neocortex shared across mammals (Emera et al., 2016).

## 4 Discussion

We presented a profile of PSI elements distributed across the genome for the first time. These PSI elements jointly covered only 0.81% of the genome space but colocalized with GWAS common variants in 2.5% of the non-overlapping PSI regions derived. Disease



variants appeared to be more likely to locate in a PSI region than quantitative trait variants, particularly in those without obvious regulatory roles (Figure 1). We further showed in the *LRP4* exemplar region the colocalization of PSI elements with common and rare variants as well as additional functional elements (Figure 2) signaling complex regulatory mechanisms. These results together render PSI elements as a valuable source to inform genetic studies of complex diseases.

It is not unexpected that majority of the mapped PSI regions are regulatory because GWAS variants tend to appear in regulatory regions (Edwards et al., 2013; Buniello et al., 2019). However, since PSI regulatory elements conveying evolution perspectives, they can bring new insights into complex regulation mechanisms as shown in the *LRP4* example here and previous examples of *SLC2A9* regulating urate levels (Takei et al., 2021) and *CACNA1C* regulating Bipolar Disorder and Schizophrenia (Song et al., 2018). On the other hand, non-regulatory PSI regions are clearly involved in various phenotypes particularly diseases. Considering the downstream *hsInsert* in the *LRP4* example (Figure 2), within it we did find rs186930464, instead of being fixed in human genome (Kronenberg et al., 2018), with varied allele frequencies in the 1,000 Genomes project and strong eQTLs and sQTLs in multiple tissues (<https://gtexportal.org/home/snp/rs186930464>) (Supplementary Tables S4, S5). Integrating existing functional information of its overlapped regulatory elements could better our understanding of the *hsInsert* and other PSI elements that were often derived from small samples with limited replication. Further functional investigation is needed to understand how these “non-regulatory” PSI elements work in shaping disease phenotypes.

The observed colocalization of PSI regions with GWAS common variants can be regarded as new evidence supporting the hypothesis of speciation characterized by substantial rewiring of the regulatory circuitry

of functional genes (Castelijns et al., 2020), suggesting that every human trait would have some human-specific characteristics, as previously showed in immune and metabolic conditions (O’Bleness et al., 2012). Since PSI elements are shared across human populations, they are likely associated with common diseases but unlikely with rare diseases.

This pilot study is limited by incomprehensive inclusion of PSI elements and incomprehensive analysis scenarios. Caution is therefore recommended when interpreting the results. Further investigations are warranted to comprehensively explore PSI values in informing genetic studies of complex diseases.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## Ethics statement

Ethical review and approval were not required for this study that used published summary data only.

## Author contributions

W-HW conceived the idea of the study. W-HW and HG contributed to the design of the study and the acquisition, analysis or interpretation of the data. W-HW performed the statistical analysis and prepared the first draft of the manuscript. All authors contributed to the content of the paper, reviewed and

approved the final version and had responsibility for the decision to submit for publication.

## Funding

This research was funded by Health Research Council NZ (HRC 17/288). W-HW was partially funded by Cure Kids NZ and the University of Otago.

## Acknowledgments

We are grateful for support from Professor Stephen Robertson and the Clinical Genetics Group in the University of Otago. We thank Professor Chris Haley of University of Edinburgh and reviewers for their valuable comments. We are grateful for access to RESTRICTED data provided by consortia ENIGMA and CHARGE, and for publicly available data provided by various consortia. We thank Grant Taylor, Robert Densie, Anna Young, Jarren Nelson and Tim Young of University of Otago for IT support.

## References

- Boyd, J. L., Skove, S. L., Rouanet, J. P., Pilaz, L. J., Bepler, T., Gordan, R., et al. (2015). Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. *Curr. Biol.* 25 (6), 772–779. doi:10.1016/j.cub.2015.01.041
- Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47 (D1), D1005–D1012. doi:10.1093/nar/gky1120
- Castelijns, B., Baak, M. L., Timpanaro, I. S., Wiggers, C. R. M., Vermunt, M. W., Shang, P., et al. (2020). Hominin-specific regulatory elements selectively emerged in oligodendrocytes and are disrupted in autism patients. *Nat. Commun.* 11 (1), 301. doi:10.1038/s41467-019-14269-w
- Doan, R. N., Bae, B. I., Cubelos, B., Chang, C., Hossain, A. A., Al-Saad, S., et al. (2016). Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* 167 (2), 341–354.e12. doi:10.1016/j.cell.2016.08.071
- Edwards, S. L., Beesley, J., French, J. D., and Dunning, A. M. (2013). Beyond GWASs: Illuminating the dark road from association to function. *Am. J. Hum. Genet.* 93 (5), 779–797. doi:10.1016/j.ajhg.2013.10.012
- Emera, D., Yin, J., Reilly, S. K., Gockley, J., and Noonan, J. P. (2016). Origin and evolution of developmental enhancers in the mammalian neocortex. *Proc. Natl. Acad. Sci. U. S. A.* 113 (19), E2617–E2626. doi:10.1073/pnas.1603718113
- Gittelman, R. M., Hun, E., Ay, F., Madeoy, J., Pennacchio, L., Noble, W. S., et al. (2015). Comprehensive identification and analysis of human accelerated regulatory DNA. *Genome Res.* 25 (9), 1245–1255. doi:10.1101/gr.192591.115
- Juan, D., Santpere, G., Kelley, J. L., Cornejo, O. E., and Marques-Bonet, T. (2023). Current advances in primate genomics: Novel approaches for understanding evolution and disease. *Nat. Rev. Genet.* 2023. doi:10.1038/s41576-022-00554-w
- Kronenberg, Z. N., Fiddes, I. T., Gordon, D., Murali, S., Cantsilieris, S., Meyerson, O. S., et al. (2018). High-resolution comparative analysis of great ape genomes. *Science* 360 (6393), eaar6343. doi:10.1126/science.aar6343
- Levchenko, A., Kanapin, A., Samsonova, A., and Gainetdinov, R. R. (2018). Human accelerated regions and other human-specific sequence variations in the context of evolution and their relevance for brain development. *Genome Biol. Evol.* 10 (1), 166–188. doi:10.1093/gbe/evx240
- Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., et al. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* 45 (6), 580–585. doi:10.1038/ng.2653
- Lu, J., Dalbeth, N., Yin, H., Li, C., Merriman, T. R., and Wei, W. H. (2019). Mouse models for human hyperuricaemia: A critical review. *Nat. Rev. Rheumatol.* 15 (7), 413–426. doi:10.1038/s41584-019-0222-x
- McLean, C. Y., Reno, P. L., Pollen, A. A., Bassan, A. I., Capellini, T. D., Guenther, C., et al. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* 471 (7337), 216–219. doi:10.1038/nature09774
- O'Bleness, M., Searles, V. B., Varki, A., Gagneux, P., and Sikela, J. M. (2012). Evolution of genetic and genomic features unique to the human lineage. *Nat. Rev. Genet.* 13 (12), 853–866. doi:10.1038/nrg3336
- Parikshak, N. N., Swarup, V., Belgard, T. G., Irimia, M., Ramaswami, G., Gandal, M. J., et al. (2016). Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* 540 (7633), 423–427. doi:10.1038/nature20612
- Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M. A., Coppens, S., Pedersen, J. S., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443 (7108), 167–172. doi:10.1038/nature05113
- Pollen, A. A., Kilik, U., Lowe, C. B., and Camp, J. G. (2023). Human-specific genetics: New tools to explore the molecular and cellular basis of human evolution. *Nat. Rev. Genet.* 2023, 1–25. doi:10.1038/s41576-022-00568-4
- Prabhakar, S., Visel, A., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., et al. (2008). Human-specific gain of function in a developmental enhancer. *Science* 321 (5894), 1346–1350. doi:10.1126/science.1159974
- Prescott, S. L., Srinivasan, R., Marchetto, M. C., Grishina, I., Narvaiza, I., Selleri, L., et al. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* 163 (1), 68–83. doi:10.1016/j.cell.2015.08.036
- R Core Team. R (2018). *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reilly, S. K., Yin, J., Ayoub, A. E., Emera, D., Leng, J., Cotney, J., et al. (2015). Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Sci. (New York, NY)* 347 (6226), 1155–1159. doi:10.1126/science.1260943
- Short, P. J., McRae, J. F., Gallone, G., Sifrim, A., Won, H., Geschwind, D. H., et al. (2018). De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature* 555 (7698), 611–616. doi:10.1038/nature25983
- Song, J. H. T., Lowe, C. B., and Kingsley, D. M. (2018). Characterization of a human-specific tandem repeat associated with bipolar disorder and Schizophrenia. *Am. J. Hum. Genet.* 103 (3), 421–430. doi:10.1016/j.ajhg.2018.07.011
- Takei, R., Cadzow, M., Markie, D., Bixley, M., Phipps-Green, A., Major, T. J., et al. (2021). Trans-ancestral dissection of urate- and gout-associated major loci SLC2A9 and ABCG2 reveals primate-specific regulatory effects. *J. Hum. Genet.* 66 (2), 161–169. doi:10.1038/s10038-020-0821-z
- Tay, S. K., Blythe, J., and Lipovich, L. (2009). Global discovery of primate-specific genes in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* 106 (29), 12019–12024. doi:10.1073/pnas.0904569106
- Vermunt, M. W., Tan, S. C., Castelijns, B., Geeven, G., Reinink, P., de Bruijn, E., et al. (2016). Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nat. Neurosci.* 19 (3), 494–503. doi:10.1038/nn.4229
- Won, H., de la Torre-Ubieta, L., Stein, J. L., Parikshak, N. N., Huang, J., Opland, C. K., et al. (2016). Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* 538 (7626), 523–527. doi:10.1038/nature19847
- Won, H., Huang, J., Opland, C. K., Hartl, C. L., and Geschwind, D. H. (2019). Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. *Nat. Commun.* 10 (1), 2396. doi:10.1038/s41467-019-10248-3

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2023.1161167/full#supplementary-material>