



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Human Genetics Seen Through An Evolutionary Lens

**Citation for published version:**

Ponting, CP 2023, 'Human Genetics Seen Through An Evolutionary Lens', *Cell Genomics*, vol. 3, 100323.  
<https://doi.org/10.1016/j.xgen.2023.100323>

**Digital Object Identifier (DOI):**

[10.1016/j.xgen.2023.100323](https://doi.org/10.1016/j.xgen.2023.100323)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Cell Genomics

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## Human Genetics Seen Through An Evolutionary Lens

Chris P Ponting, MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK.

Correspondence: [Chris.Ponting@ed.ac.uk](mailto:Chris.Ponting@ed.ac.uk).

### Abstract

Most DNA bases crucial for species perpetuation are marked by a dearth of sequence change among species related over long evolutionary time. Recently, Christmas et al. <sup>1</sup> and Sullivan et al. <sup>2</sup> cast light on human DNA and its variants by comparisons with 239 other mammalian species' genomes.

### Main body

Human DNA, like any other genome, has been shaped by a strong survivorship bias. Bases of importance to species propagation have been retained preferentially, whilst others of less or no such importance have been mutated across tens or hundreds of millions of years of evolution. This explains why the degree of evolutionary conservation has long been used as a qualitative rule-of-thumb when assessing the importance of a gene or its regulatory elements. To provide a more detailed quantitation of evolutionary conservation, however, the genomes of 240 placental mammal species needed to be sequenced and multiply aligned <sup>3</sup>. Species that are less closely related to humans, for example birds or fish, were not chosen for this purpose despite providing the long evolutionary divergence needed to pinpoint highly conserved functional DNA. This is because mammalian phyla also carry DNA that has acquired functionality, or else has dispensed with it, only recently ( $\sim 10^7$  years) in a lineage-specific manner <sup>4</sup>. Comparing human with bird or fish DNA would focus on deeply conserved functions, but would often miss mammalian-specific innovations. In a recent paper, Christmas et al. <sup>1</sup> further analyse these 240 genomes, finding that about 10% of the human genome is conserved across these mammals, about a third of which can be pinpointed at single base resolution. This consortium's second paper (Sullivan et al. <sup>2</sup>) then examined these conserved single bases, finding that variation at these sites occurs infrequently in our population, as expected from their deleteriousness.

Christmas et al. estimated that at least 11.6% (332Mb of 2.85Gb) of the human genome has been constrained over mammalian evolution. This is an important number to estimate because it is also the approximate fraction of all newly arising DNA variants that affect functional sequence. The estimate is a third-higher than a previous prediction <sup>4</sup> (253 Mb; 95% CI 220–286 Mb) which also took account of mammalian lineage-specific constraint. Christmas et al.'s estimates of constrained sequence varied widely across four other focal species (239-367Mb) leading them to conclude that their method cannot determine whether the true extent of constraint varies among mammals. Conservation across mammals matched long-held notions of importance: protein-coding sequence, including their start and stop codons, were most enriched in conserved bases, followed by genes' untranslated regions and promoters; also, developmental genes were most conserved, whereas environmental response genes were least conserved.

Impressively, the consortium could pinpoint a third of all constrained sequence (101Mb) at single base resolution. Half of this amount lies outside of protein-coding sequence and is functionally unannotated, creating substantial opportunities for large-scale reverse genetics experiments. To gain insights into the molecular function of this vast genomic terra incognita the Zoonomia team intersected constrained sequence with 15.6 million transcription factor (TF) binding sites. The great majority (88%) of these sites, however, lacked conservation and they found only a handful of TFs whose conserved binding sites outnumbered their non-conserved sites. This indicates that binding of

TFs only rarely regulates transcription, as previously proposed <sup>5</sup>. As Doolittle et al.<sup>6</sup> suggested, only sites whose DNA variants are under selection should be deemed functional.

Sullivan et al.<sup>2</sup> compared constraint that shaped mammalian genomes over long evolutionary time, with constraint applied across more recent human evolution. As expected, they found human genetic variation at mammal-constrained positions being preferentially purged from our population. Over a third of all human genes lacked any common or low frequency variants at mammal-conserved sites, including many involved in fundamental processes such as in embryogenesis. Genes with greater human variation at mammal-conserved sites tended to be more involved in environmental response, including immunity. Human long noncoding RNA promoters and exons lacked mammal-conserved sites almost entirely, as reported previously <sup>7</sup>. Overall, common and low frequency human variants at mammalian-conserved sites were more likely to be pathogenic and to have a larger influence on human traits. These observations will help improve the fine-mapping of disease-causal variants and estimating the inherited liability of genetic disease.

Zoonomia has gifted science an important data resource whose value will steadily rise from future use. The project recapitulates past findings gleaned from much smaller data sets, principally that the less variable a DNA site is among humans, the stronger is its constraint across mammals. The project should also be viewed as signposting future genetics and genomics research opportunities. Seen with a critical eye, Zoonomia's broad sampling of eutherian phyla will result in diminishing returns from future genome sequencing of additional species (the genomes of about 5,900 placental mammal species remain to be sequenced <sup>1</sup>). Also, their genome analyses make clear that few inferences can be drawn about each species' distinctive traits from its genome sequence alone. Despite their claims <sup>1</sup>, sadly it is unlikely that Zoonomia will substantively protect future biodiversity. Rather the project's data may ultimately be viewed as cataloguing the rich inheritance of a once widely successful animal clade.

Viewed more positively, future opportunities now present themselves. The first of these is more accurate inference of constrained sequence. Zoonomia took advantage of methods that did not account for known mutation rate variation at the ~10Mb scale <sup>8</sup>. Future methods that account for chromosomal location- and phyla-dependent variation may help pinpoint lineage-specific functional sequence, which is estimated to contribute about half of all constrained DNA in placental mammals <sup>4</sup>. Second, Zoonomia alignments may reveal correlated evolutionary sequence changes in protein-coding or non-coding sequence, each indicative of multiple compensatory mutations and potentially revelatory of mechanism. Finally, Zoonomia has shone a bright light on >400,000 conserved non-coding elements (median size 20 nucleotides) that remain unannotated and unexplained. This re-emphasises how incompletely regulatory function has been interrogated in cell-based assays to date. Particularly under-represented will be noncoding regulatory function in rare cell types, or involved in early developmental processes or responding to environmental conditions difficult to replicate in laboratory conditions. The greatest challenge now is to couple Zoonomia's deep evolutionary information with novel assays in order to reveal regulatory molecular mechanisms.

### **Acknowledgements**

CPP group research is funded by the MRC (MC\_PC\_20005 and MC\_UU\_12008/1).

### **References**

1. Christmas, M.J., Kaplow, I.M., Genereux, D.P., Dong, M.X., Hughes, G.M., Li, X., Sullivan, P.F., Hindle, A.G., Andrews, G., Armstrong, J.C. et al. (2023). Evolutionary constraint and innovation across hundreds of placental mammals. *Science*, in press.

2. Sullivan, P.F., Meadows, J.R.S., Gazal, S., Phan, B.N., Li, X., Genereux, D.P., Dong, M.X., Bianchi, M., Andrews, G., Saktikumar, S. et al. (2023). Leveraging Base Pair Mammalian Constraint to Understand Genetic Variation and Human Disease. *Science*, in press.
3. Zoonomia Consortium. (2020). A comparative genomics multitool for scientific discovery and conservation. *Nature* 587, 240-245.
4. Rands, C.M., Meader, S., Ponting, C.P., and Lunter, G. (2014) 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genet.* 10, e1004525.
5. Cusanovich, D.A., Pavlovic, B., Pritchard, J.K., and Gilad, Y. (2014) The Functional Consequences of Variation in Transcription Factor Binding. *PLoS Genet.* 10, e1004226.
6. Doolittle, W.F., Brunet, T.D.P., Linquist, S., and Gregory, T.R. (2014) Distinguishing between “Function” and “Effect” in Genome Biology. *Genome Biol. Evol.* 6, 1234-1237.
7. Haerty W., and Ponting, C.P. (2014) No gene in the genome makes sense except in the light of evolution. *Annu Rev Genomics Hum Genet.* 15, 71-92
8. Hodgkinson, A., and Eyre-Walker, A. (2011) Variation in the mutation rate across mammalian genomes. *Nat. Rev. Genet.* 12, 756-766.