JASs Journal of Anthropological Sciences Open Access
forum Vol. 99 (2021), pp. 171 - 174 doi. 10.4436/jass.99006

Towards a functional understanding of adaptive phenotypes in humans

Maxime Rotival[1] & Lluis Quintana-Murci[1,2]

1) Human Evolutionary Genetics Unit, Institut Pasteur, UMR 2000, CNRS, Paris 75015, France
2) Chair of Human Genomics and Evolution, Collège de France, Paris, 75005, France
e-mail: quintana@pasteur.fr

Over the last decades, anthropological and population genetics studies have provided robust evidence of genetic adaptation of human populations to their local environments (Fan et al. 2016). Iconic examples include selection for light skin color in West Eurasia, lactose tolerance in agricultural communities around the world, and malaria resistance in populations living in subtropical areas endemic from *Plasmodium vivax* and *Plasmodium falciparum,* the causative agents of malaria in humans. While the characterization of human genetic diversity at the worldwide scale and the growing understanding of how selection operates, through selective sweeps or polygenic adaptation, have allowed to drastically expand the number of candidate selected loci, the identification of genes and functions genuinely targeted by natural selection and the underlying causal variants is still lagging behind.

Three main factors can explain such knowledge gap. First, despite considerable efforts have been done (Akbari et al. 2018; Sugden et al. 2018), pinpointing the adaptive variants based on population genetics approaches alone has proved a daunting task. Indeed, *genetic hitchhiking* causes neutral alleles that segregate on the same haplotype as the beneficial mutation to increase in frequency concomitantly, leading to signatures of positive selection similar to those observed at the causal variant. Second, when the adaptive variant is non-coding, long-range interactions between enhancers and their target genes, and tissue-dependent effects on gene expression – where the same enhancer activates different genes depending on the tissue – can further obscure the identity of the genes under selection. Third, even when the target gene is identified, pleiotropy (i.e., when a single gene is associated with multiple phenotypes) often complicates the identification of the true adaptive phenotype.

There is thus a clear need to characterize the functional consequences of the signals of genetic adaptation detected in the genomes of modern humans. This can be achieved through a combination of approaches that we discuss below, including (i) association studies of molecular and clinical phenotypes to assess the context-specific effects of adaptive loci on protein function and their impact on downstream-phenotypes; (ii) massively parallel assays to study the functional impact of specific candidate variants under selection; and (iii) genome-wide approaches to identify the specific phenotypes that have been the targets of selection.

A first step towards the understanding of the functional significance of the adaptive events detected by anthropological studies is the systematic functional annotation of the selected variants and haplotypes. The increasing availability of summary statistics from genome-wide association studies (GWAS) and the collection of extensive genetic and phenotypic data in large cohorts, such as the UK biobank (Bycroft et al. 2018), are allowing to perform phenome-wide association studies (PheWAS), which assess the effect of a single genetic variant on a large range of phenotypes. Although such studies are essential to unbiasedly assess the phenotypic effects of adaptive loci, they provide limited information on the underlying mechanisms. In this context, studies of expression quantitative trait loci (eQTL), which measure the effects of genetic variants on

gene expression, have proven very valuable. A large number of eQTL databases exists, allowing to study the molecular effects of genetic variants across multiple tissues (GTEx Consortium 2020), immune cell types (Schmiedel et al. 2018), or in response to external stimuli such as pathogens (Quach et al. 2016). Yet, the majority of these studies are heavily biased towards European-ancestry populations, and there is an urgent need to expand them to a wider range of ancestries, in particular to neglected populations (Sirugo et al. 2019). Likewise, eQTL studies need to include a larger range of stimuli and cellular states to better understand how genetics controls gene expression in the presence of other immune ligands, hormones, drugs, vitamins or metabolites.

The ongoing revolution brought by single-cell approaches should also allow to better apprehend how genetic variants act in a context-specific manner. For example, single-cell RNA-sequencing of induced pluripotent stems cells allows capturing the transient effect of genetic variants that influence expression during differentiation and control cellular fate (Cuomo et al. 2020). This approach will facilitate the study of the genetic basis of human traits with a developmental or hematopoietic origin, such as stature, brain-structure, or blood cell counts. Such a technology will also be extremely valuable in the context of immune response to viruses, where the genetic control of gene expression may strongly differ between infected cells and those exposed to the virus but resistant to infection (i.e., bystander cells).

To better understand the functional impact of adaptive variants on phenotype variation, further effort should be also placed into how they alter protein function. Beyond their effects on transcription, non-coding variants can also alter mRNA splicing, degradation, or translation. For instance, adaptive introgression from Neanderthals at the antiviral *OAS* locus has been associated with altered splicing of the terminal exons of *OAS1*, leading to the formation of an isoform that alters cellular localization and increases antiviral activity (Sams et al. 2016). Hence it is crucial to routinely complement standard eQTL analyses with analyses of splicing QTLs or even protein QTLs. The integration of these findings across a large number of tissues and conditions should allow to fully characterize the specific contexts in which selected haplotypes alter the function of nearby genes.

While association studies are a powerful tool to characterize the phenotypic effects of genes presenting signals of selection, they are highly sensitive to linkage disequilibrium, and often fall short of identifying the specific functional variants. Solving this issue requires to separately test the effect of each variant using complex functional assays, such as luciferase reporter assays or knock-out experiments. The recent advent of massively parallel assays is now allowing to perform these experiments at an unprecedented scale. For example, pooled CRISPR screens combine the power of targeted DNA editing with single-cell sequencing to simultaneously alter thousands of regulatory elements and monitor their effect on gene expression (Gasperini et al. 2019). Notably, a recent study has applied massively parallel reporter assays to characterize the regulatory effects across three tissues (embryonic stem cells, brain, and osteoblasts) of over 14,000 alleles that have reached fixation in modern humans, but are absent from the genome of other hominins such as Neanderthals or Denisovans (Weiss et al. 2021). This study has revealed an enrichment of functional variants near genes controlling vocal tract and brain anatomy, demonstrating the power of this approach to unravel the specific genetic variants that have shaped the evolution of modern humans.

Even when a selection signal can be narrowed down to a single genetic variant, it can still regulate multiple phenotypes, given that pleiotropy appears to be the rule rather than the exception. For example, the rs4988235-T allele at the *LCT* locus, known to be associated with lactase persistence in adulthood, is associated with over 111 additional phenotypes in the UK biobank cohort (Bycroft et al. 2018). While some of these phenotypes may result from the lactase persistency effect, such as altered dietary habits and increased weight and fat percentage, others

appear to be unrelated, including decreased ease of skin tanning, higher blood pressure, or altered social behavior. Thus, the pervasive pleiotropy observed in the genome should incite caution when concluding on the specific phenotypes that are targeted by selection.

Recent studies of the genetic architecture of human traits have shown that most complex traits – such as those that determine survival and reproductive success – are highly polygenic, with the most constrained traits usually being the most polygenic. In this context, our best chance at deciphering truly adaptive phenotypes might be to adopt to a genome-wide view of selection, and seek for an accumulation of selection signals across genes contributing to the same phenotype. For example, an ancient DNA study of 170 skin pigmentation loci has revealed strong directional selection for lighter skin color in Eurasia over the last 40,000 years, with the strongest selection acting on a subset of 10-20 loci (Ju and Mathieson 2021). This finding reinforces the notion that selection often acts through multiple targets to reach the advantageous phenotype. Furthermore, comparing the number of loci associated to the same phenotype can be used to separate primary selected phenotypes from hitchhiking phenotypes. For example, in a population genetic study of Central-African rainforest hunter-gatherers, we have recently shown an enrichment of selection signals near genes associated with both height and reproductive age (Lopez et al. 2019). Yet, the signals at reproduction-related traits vanished when excluding height-associated loci from the analyses, whereas the removal of reproduction-related loci did not affect the enrichment of selection signals at height-associated genes. This supports that height is the truly adaptive phenotype that, due to pleiotropy, affects reproductive age in this hunter-gatherer group. This case of adaptation to the extreme conditions of the rainforest illustrates how adopting a genome-wide approach can help deciphering the nature of adaptive phenotypes.

Together, while the road towards a better understanding of the effects of natural selection on human phenotypes is steep, a promising path lies ahead by combining the mapping of molecular quantitative trait loci across multiple tissues and conditions, with the functional dissection of selected loci by massively parallel assays and the genome-wide screening of phenotypes and biological functions associated with adaptive loci. In doing so, the coming years will bring novel insights, and possibly some surprises, into the mechanisms through which humans have biologically adapted to the multiple environments they have encountered over the last 60,000 years of evolution.

Acknowledgements

The laboratory of Human Evolutionary Genetics is supported by the Institut Pasteur, the Collège de France, the Centre Nationale de la Recherche Scientifique (CNRS), the Agence Nationale de la Recherche (ANR) grants LIFECHANGE (ANR-17- CE12-0018-02), CNSVIRGEN (ANR-19-CE15-0009-02) and COVID-19-POPCELL (ANR-21-CO14-003-01), the French Government's Investissement d'Avenir program, Laboratoires d'Excellence 'Integrative Biology of Emerging Infectious Diseases' (ANR-10- LABX-62-IBEID) and 'Milieu Intérieur' (ANR-10-LABX-69-01), the Fondation pour la Recherche Médicale (Equipe FRM DEQ20180339214), the Fondation Allianz-Institut de France, and the Fondation de France (n°00106080).

References

Akbari A, Vitti JJ, Iranmehr A, Bakhtiari M, et al (2018) Identifying the favored mutation in a positive selective sweep. Nat Methods 15:279-282. https://doi.org/10.1038/nmeth.4606

Bycroft C, Freeman C, Petkova D, et al (2018) The UK Biobank resource with deep phenotyping and genomic data. Nature 562:203-209. https://doi.org/10.1038/s41586-018-0579-z

GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369:1318-1330. https://doi.org/10.1126/science.aaz1776

Cuomo ASE, Seaton DD, McCarthy DJ, et al (2020) Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. Nat Commun 11:810. https://doi.org/10.1038/s41467-020-14457-z

Fan S, Hansen ME, Lo Y, et al (2016) Going global by adapting local: A review of recent human adaptation. Science 354:54-59. https://doi.org/10.1126/science.aaf5098

Gasperini M, Hill AJ, McFaline-Figueroa JL, et al (2019) A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. Cell 176:377-390 e19. https://doi.org/10.1016/j.cell.2018.11.029

Ju D, Mathieson I (2021) The evolution of skin pigmentation-associated variation in West Eurasia. Proc Natl Acad Sci USA 118:e2009227118. https://doi.org/10.1073/pnas.2009227118

Lopez M, Choin J, Sikora M, et al (2019) Genomic Evidence for Local Adaptation of Hunter-Gatherers to the African Rainforest. Curr Biol 29:2926-2935e4. https://doi.org/10.1016/j.cub.2019.07.013

Quach H, Rotival M, Pothlichet J, et al (2016) Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell 167:643-656 e17. https://doi.org/10.1016/j.cell.2016.09.024

Sams AJ, Dumaine A, Nedelec Y, et al (2016) Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. Genome Biol 17:246. https://doi.org/10.1186/s13059-016-1098-6

Schmiedel BJ, Singh D, Madrigal A, et al (2018) Impact of genetic polymorphisms on human immune cell gene expression. Cell 175:1701-1715 e16. https://doi.org/10.1016/j.cell.2018.10.022

Sirugo G, Williams SM, Tishkoff SA (2019) The Missing Diversity in Human Genetic Studies. Cell 177:26-31. https://doi.org/10.1016/j.cell.2019.02.048

Sugden LA, Atkinson EG, Fischer AP, et al (2018) Localization of adaptive variants in human genomes using averaged one-dependence estimation. Nat Commun 9:703. https://doi.org/10.1038/s41467-018-03100-7

Weiss CV, Harshman L, Inoue F, et al (2021) The cis-regulatory effects of modern human-specific variants. Elife 10:e63713. https://doi.org/10.7554/eLife.63713

This work is distributed under the terms of a Creative Commons Attribution-NonCommercial 4.0 Unported License http://creativecommons.org/licenses/by-nc/4.0/