Human Genetics (2021) 140:241–276
https://doi.org/10.1007/s00439-020-02206-7

REVIEW

Identifying adaptive alleles in the human genome: from selection mapping to functional validation

Elizabeth A. Werren[1,2] · Obed Garcia[2] · Abigail W. Bigham[3]

Received: 19 March 2020 / Accepted: 7 July 2020 / Published online: 29 July 2020

Abstract

The suite of phenotypic diversity across geographically distributed human populations is the outcome of genetic drift, gene flow, and natural selection throughout human evolution. Human genetic variation underlying local biological adaptations to selective pressures is incompletely characterized. With the emergence of population genetics modeling of large-scale genomic data derived from diverse populations, scientists are able to map signatures of natural selection in the genome in a process known as selection mapping. Inferred selection signals further can be used to identify candidate functional alleles that underlie putative adaptive phenotypes. Phenotypic association, fine mapping, and functional experiments facilitate the identification of candidate adaptive alleles. Functional investigation of candidate adaptive variation using novel techniques in molecular biology is slowly beginning to unravel how selection signals translate to changes in biology that underlie the phenotypic spectrum of our species. In addition to informing evolutionary hypotheses of adaptation, the discovery and functional annotation of adaptive alleles also may be of clinical significance. While selection mapping efforts in non-European populations are growing, there remains a stark under-representation of diverse human populations in current public genomic databases, of both clinical and non-clinical cohorts. This lack of inclusion limits the study of human biological variation. Identifying and functionally validating candidate adaptive alleles in more global populations is necessary for understanding basic human biology and human disease.

Introduction

The evolution of modern humans involved extensive migration to diverse geographic regions with shifting environmental conditions. Changes in climate, diet, and novel exposures to pathogens coincided with transitions in sociocultural organization (Laland et al. 2010). During these shifts, changes in the genome that initially arose as stochastic events within individuals of populations were subjected to evolutionary forces over generational time. These forces of random genetic drift, gene flow, and natural selection have together shaped the genetic diversity of our species. Evidence supports that a large portion of this variation has evolved neutrally (Harris 2018; Kimura 1968, 1983). Nonetheless, environmental and cultural factors jointly have imposed selective pressures on the human genome, contributing to extant patterns of human genetic diversity (Feldman and Cavalli-Sforza 1976). Local adaptation as result of natural selection contributes to the phenotypic variation observed across human populations (Bamshad and Wooding 2003; Vitti et al. 2013). Deciphering which genetic changes from natural selection underlie adaptive traits in humans is an ongoing area of investigation in evolutionary biology. Additionally, current environments may no longer be suitable for formerly beneficial alleles, resulting in a gene–environment mismatch underlying human disease (Gluckman et al. 2009; Nesse et al. 2010; Nesse and Stearns 2008; Williams and Nesse 1991). Determining the genetic basis of human adaptation therefore has broad implications for understanding human biology, health, and disease.

Natural selection can be described in three general categories: positive directional selection, balancing selection, and negative purifying selection (Fig. 1a). Each mode of natural selection will influence the alleles in a population in different ways (Gillespie 1994). Positive selection favors

✉ Abigail W. Bigham
awbigham@ucla.edu

[1] Department of Human Genetics, The University of Michigan, Ann Arbor, MI, USA

[2] Department of Anthropology, The University of Michigan, Ann Arbor, MI, USA

[3] Department of Anthropology, University of California Los Angeles, 341 Haines Hall, Los Angeles, CA 90095, USA

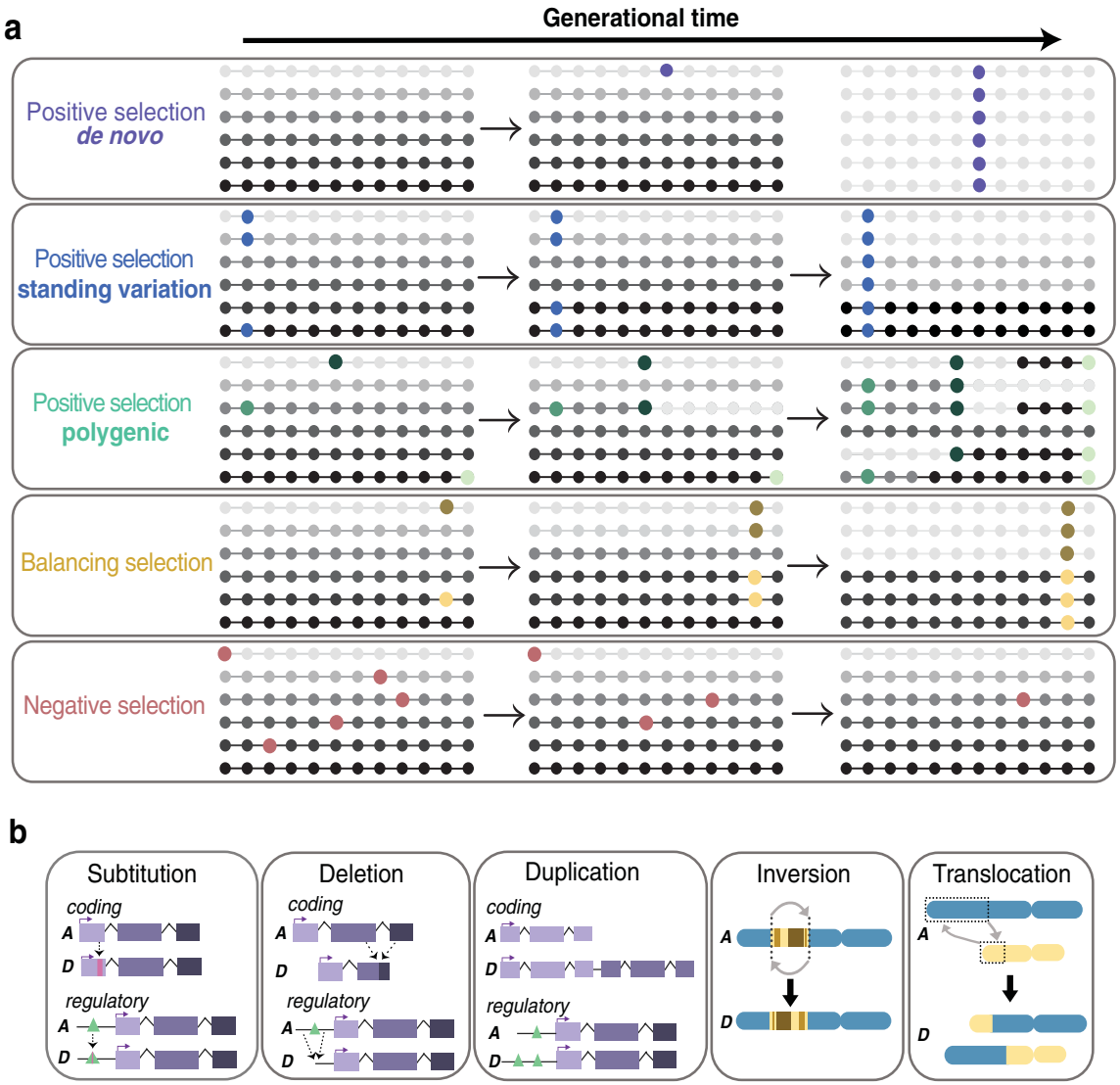

Fig. 1 Natural selection influences genomic diversity. **a** Theoretical basis for population genetics models of natural selection. Each dotted strand represents a haplotype. Alleles that increase or decrease in frequency over time due to selection are represented as dots in purple (positive selection on de novo variants), blue (positive selection on standing variation), green (polygenic positive selection), yellow (balancing selection), red (negative selection), and shades of grey (linked variation to causal allele). **b** Types of adaptive genetic variants that can be acted upon by natural selection. *A*, ancestral; *D*, derived

the fixation of an advantageous allele that increases the evolutionary fitness of its carrier. Balancing selection acts to maintain frequencies of multiple alleles at a locus for fitness benefit. Purifying selection, also known as background selection, removes deleterious alleles that reduce the fitness of the carrier from the population (Charlesworth et al. 1993). Purifying selection is essential for the evolutionary conservation of biological function (Cvijovic et al. 2018). The effects of strong purifying selection on genomic diversity have been characterized in several organisms (Begun and Aquadro 1992; Cvijovic et al. 2018; McVicker et al. 2009). Here, we focus on the latter two forms of natural selection, positive and balancing, as these forms lead to the fixation or maintenance of adaptive alleles in a population over time (Maynard-Smith and Haigh 1974). Furthermore, these evolutionarily favored genetic changes can be identified through a process called selection mapping and functionally assessed for relevance to proposed adaptive biologies.

Selection mapping refers to the process of using population genetics methods to detect genomic regions, genes, or variants that have been acted on by natural selection (Wisser et al. 2008). It is founded on the principle that phenotypic change in a population is accompanied by changes in allele frequencies (Fumagalli et al. 2010; Vitti et al. 2013; Wisser et al. 2008). Population-specific changes in allele frequency due to natural selection can leave distinct statistical

signatures at the target loci relative to neutrally evolving loci (Fu 1995; Fu and Li 1993). Inferred signals of selection can be leveraged to identify fitness-related genomic regions, genes, and/or variants. Larger detected genomic regions can be further fine-mapped to pinpoint putative functional alleles contributing to adaptive phenotypes (Akbari et al. 2018; Schaid et al. 2018; Szpak et al. 2018). Indeed, selection mapping has been widely used by evolutionary and population geneticists to identify candidate functional alleles underlying adaptive phenotypic change in plant and animal populations, including humans (Akey et al. 2002; Matsumoto et al. 2017; Wisser et al. 2008).

This review will discuss the application of selection mapping approaches to identify genetic variation contributing to adaptive phenotypes across human populations, thus focusing on the microevolutionary scale. We begin by describing the search for candidate adaptive alleles, from the detection of selection signatures (namely positive and balancing) through genotypic-phenotypic association and fine mapping variants. We provide a review of the discovered candidate adaptive alleles in human populations followed by a summary of molecular approaches for functional validation of their biological role. While the bulk of genotyping and sequencing efforts have focused on European populations (Popejoy and Fullerton 2016; Sirugo et al. 2019), we detail several cases of selection mapping in non-European populations. Nevertheless, we acknowledge the impact of this major Eurocentric bias in the current map of natural selection in humans, and we conclude with comments on future directions and inclusivity considerations for population geneticists, human geneticists, and molecular anthropologists.

Detecting natural selection in the genome

Genetic variants that define the alleles subjected to natural selection, due to effects on biological fitness, are distributed across functional coding and non-coding genomic elements (Bamshad and Wooding 2003; Jha et al. 2015; Wooding et al. 2002; Zhao et al. 2000). The types of genetic variants targeted by natural selection include nucleotide substitution, insertion, deletion, duplication, inversion, and translocation (Fig. 1b). Natural selection can act on either de novo genetic variant(s) or standing genetic variation, leaving behind distinct signatures in the genome (Barrett and Schluter 2008; Hermisson and Pennings 2005; Przeworski et al. 2005). For selection on de novo variation, a novel variant arises in a population in which a selective pressure is present (Peter et al. 2012). For selection to act, this new variant will confer either an adaptive advantage or disadvantage to the existing pressure. Selection on standing variation may occur when a population faces an environmental change or range expansion that results in a new selective pressure (Przeworski et al. 2005). Thus, standing variation is preexisting variation in the ancestral population at the time the environmental conditions for selection appear (Barrett and Schluter 2008). Depending on the mode of natural selection, genomic diversity will be shaped differently in a population over time.

Distinguishing selection on de novo variants from standing variation

The structure of haplotypes—sets of alleles (or genetic variants) that are inherited together on a single chromosome in an individual—in a population offer clues into how natural selection has acted on the genome (Bamshad and Wooding 2003; Nordborg and Tavare 2002; Sabeti et al. 2002b). In particular, the heterogeneity of haplotype backgrounds that comprise the allele under natural selection, as well as the changes in allele frequency, help to distinguish between signals of selection from standing variation versus de novo variation (Barrett and Schluter 2008; Peter et al. 2012). If a selective pressure is present before a favorable variant originates, as in the case of selection on a de novo variant, then the new (or derived) allele will come to exist on a single, distinct haplotype (Fig. 1a) in the study population, and it will be absent in the ancestral population, where only the ancestral allele will be found (Nielsen et al. 2007; Peter et al. 2012). If the derived allele reaches fixation rapidly from positive selection, haplotype homozygosity will be extensive around the selected allele before mutation and recombination act to break it down (Nielsen et al. 2007; Peter et al. 2012). In other words, the linkage disequilibrium (LD)—the non-random association of alleles in haplotypes—will be high as a result of genetic hitchhiking of neutral alleles linked to the adaptive allele during its rise in frequency (Maynard-Smith and Haigh 1974; Przeworski 2002). This is referred to as a 'hard' selective sweep. Effective population size and the effect size of the adaptive allele contribute to its fixation time, with shorter time to fixation in cases of small effective population size and large effect alleles (Pritchard et al. 2010). Once a derived allele has swept to high frequency, recombination acts to break up linkage with neutral sites resulting in an excess of both intermediate and high frequency alleles (Barrett and Schluter 2008; Fay and Wu 2000). However, if the selective sweep is incomplete, i.e., the beneficial allele is in the midst of being swept to fixation in the population, heterozygosity increases, and this is called a 'soft' sweep (Pritchard et al. 2010).

Selection also acts on standing variation. When selection favors variation already present in the population, the pre-adaptive alleles may exist on several distinct haplotypes, resulting in a greater number of linked neutral alleles at intermediate frequencies following selection (Fig. 1a) (Barrett and Schluter 2008; Peter et al. 2012; Przeworski et al. 2005). In other words, alleles that are maintained within the population as neutral or weakly functional before becoming

selectively advantageous will have a longer population history than a new adaptive allele (Barrett and Schluter 2008). This protracted population history increases the likelihood that recombination has broken up haplotype blocks containing the standing variant, placing it on different backgrounds (Barrett and Schluter 2008; Peter et al. 2012; Przeworski et al. 2005). Selection on standing variants will integrate more polymorphisms than selective sweeps of de novo variants, and this distribution across haplotypes weakens the statistical signature of selection (Barrett and Schluter 2008; Hermisson and Pennings 2005; Peter et al. 2012; Przeworski et al. 2005). Variation around the fixed adaptive allele will therefore be lower in a population following selection on de novo variation compared to selection on standing variation (Barrett and Schluter 2008). Given the predicted effects of selection on alleles and haplotypes between populations, the type of sequence information acquired (SNP microarray, targeted sequencing, whole exome, whole genome, etc.) and selection of study populations will determine which methods (Table 1) can be applied for detecting natural selection in the genome.

Haplotype tests of selection

Several extended haplotype tests for identifying selective sweeps exist and include the Long-Range Haplotype (LRH), integrated Haplotype Score (iHS), number of segregating sites by length statistic (nSL), and Cross Population Extended-Haplotype Heterozygosity (XP-EHH) (Table 1) (Sabeti et al. 2002b,2007; Voight et al. 2006). The LRH test takes into consideration both the length of extended haplotype homozygosity (EHH) and the frequency of the haplotype in the population (Sabeti et al. 2002b). This test performs well when identifying haplotypes that arose quickly and have yet to be broken down by recombination, such as for hard sweeps. A modification of this test, the whole genome long-range haplotype (WGLRH) test, has been developed to use genome-wide data to detect selection by combining LRH with derived and ancestral allele status (Zhang et al. 2006). The iHS test quantifies the strength of selection acting at a locus in a single population by calculating the integral of EHH decay at varying distances away from a core allele (Fig. 2) (Voight et al. 2006). It considers the ratio of haplotype homozygosity for haplotypes carrying the derived allele relative to those carrying the ancestral allele at the candidate locus. By doing so, iHS is sensitive to variation in recombination rates across the genome as the two alleles serve as controls for each other, canceling out heterogeneity in the estimated genetic map due to recombination coldspots and hotspots. This statistic is more powered to detect incomplete, or soft, selective sweeps (Pickrell et al. 2009; Voight et al. 2006). Similar to iHS, nSL uses the ratio of haplotype homozygosity for derived and ancestral alleles to identify positive selection in a single population (Ferrer-Admetlla et al. 2014). nSL differs from iHS in that it quantifies EHH length for a given haplotype pair based on the number of SNPs present in that region on all other haplotypes in the sample population (Ferrer-Admetlla et al. 2014; Voight et al. 2006). In doing so, a genetic map is not needed to calculate nSL, thereby further increasing sensitivity to varying recombination and mutation rates, with greater power for detecting both hard and incomplete/soft selective sweeps. XP-EHH compares haplotypes between populations to identify derived selected alleles that have recently reached or nearly reached fixation in one population yet remain polymorphic across all human populations (Sabeti et al. 2007). It performs well for detecting sweeps that have reached, or have almost reached, fixation (Pickrell et al. 2009; Sabeti et al. 2007; Voight et al. 2006). As such, it has weaker power to detect older selective sweeps that have been subjected to LD degradation by recombination over time.

Confounding variables in haplotype-based tests

Population-specific recombination rate and copy-number variants (CNVs) may confound the results of haplotype-based tests of selection and should be taken into account (Sabeti et al. 2007). Certain regions of the genome exhibit variable recombination rates between populations (Abecasis et al. 2010). A long-range haplotype signal at these loci could result from a slower recombination rate in one population relative to other populations, instead of a spike in allele frequency from natural selection (Sabeti et al. 2002b). To account for this, the level of EHH decay of non-selected alleles in the region surrounding the candidate allele can be compared to the EHH decay of those same non-selected alleles in other populations (Sabeti et al. 2002b). If the EHH decay is similar between populations, variable recombination rate can be ruled out as the source of the long-range signal, bolstering claims of selection (Sabeti et al. 2002b). CNVs can result in the increase of nearly identical sequence at a locus that will reduce heterozygosity (Lucas et al. 2019). This makes it challenging to distinguish a selective sweep from a region with high copy number variation (Iskow et al. 2012). Likewise, CNVs that lie close to a selected locus may falsely extend the signal (Sabeti et al. 2007). To account for potential confounding from CNVs, recent innovations in long-read sequencing technologies that avoid PCR amplification steps offer a platform to more accurately map and quantify CNVs (Pollard et al. 2018). Further, hybrid-sequencing approaches that combine long-read with short-read technologies that offer higher resolution and lower genotyping error rates, can afford both CNV accuracy and resolution (Kronenberg et al. 2018).

Table 1 Summary of the different types of selection detection methods outlined in this review

Type	Abbreviation	Test name	Summary	References
Haplotype-based	EHH	Extended haplotype homozygosity	The EHH calculates the decay of LD (association of alleles) surrounding a core haplotype (or locus of interest), at varying distances from the core. From this, EHH then assesses the probability that any two haplotypes in a population carrying the core haplotype are identical by descent. This is useful for detecting recent positive selection	Sabeti et al. (2002a, b)
	LRH	Long-range haplotype	The Long-Range Haplotype combines EHH with the haplotype frequency in the population to detect regions of recent positive selection, primarily hard sweeps	Sabeti et al. (2002a, b)
	WGLRH	Whole genome long-range haplotype test	The Whole Genome Long-Range Haplotype test incorporates the LRH with patterns of LD to identify regions with long haplotypes, which may be indicative of positive selection	Zhang et al. (2006)
	iHS	integrated Haplotype Score	The integrated Haplotype Score measures the ratio of EHH decay of haplotypes carrying derived alleles to those carrying ancestral alleles. This method is useful for detecting soft selective sweeps	Voight et al. (2006)
	nSL	Number of segregating sites by length statistic	The Number of Segregating Sites by Length Statistic measures the ratio of EHH for haplotypes carrying derived alleles relative to those with ancestral alleles. However, unlike iHS, it incorporates segregating sites to measure distance rather than a genetic map. nSL is useful for detecting both hard and soft selective sweeps	Ferrer-Admetlla et al. (2014)
	XP-EHH	Cross population extended haplotype homozygosity	The Cross Population Extended Haplotype Homozygosity incorporates the integral of EHH, and compares haplotypes between two populations to detect those containing nearly fixed derived alleles. This is useful for detecting hard selective sweeps	Sabeti et al. (2007)

Table 1 (continued)

Type	Abbreviation	Test name	Summary	References
Allele frequency	Tajima's *D*	Tajima's *D* statistic	Tajima's *D* measures the differences between segregating sites (Θ_W) and the average nucleotide differences (π). It can be useful for detecting departures from neutrality, which may suggest balancing or positive selection. However, it is confounded by demographic conditions, such as population structure and growth	Tajima (1989)
	Fu and Li's *D* and *F*	Fu and Li's *D* and *F* statistics	Fu and Li's *D* and *F* are based on coalescence, and compare the number of derived singleton variants to Θ_W and π, respectively. These statistics can suggest recent positive, purifying, or balancing selection. Fu and Li's *D* and *F* are more sensititve to detect selective sweeps acting on derived alleles compared to Tajima's *D*	Fu and Li (1993)
	Fay and Wu's *H*	Fay and Wu's *H* statistic	Fay and Wu's *H* looks for a high frequency of derived variants to determine whether the locus has undergone positive selection compared to neutral expectations	Fay and Wu (2000)
	F_{ST}	Wright's Fixation Index	Wright's F_{ST} is a measure of genetic differentiation between two populations. This statistic can be used to identify loci under positive or balancing selection	Wright (1950)
Phylogenetic-based	LSBL	Locus-specific branch length	The Locus Specific Branch Length incorporates F_{ST} for a three-population comparison, where distances between populations are represented by branch lengths	Shriver et al. (2004)
	PBS	Population branch statistic	The Population Branch Statistic, similar to LSBL, uses a three-population comparison based on F_{ST}, except the values are log-transformed to incorporate phylogenetics	Yi et al. (2010)

Table 1 (continued)

Type	Abbreviation	Test name	Summary	References
Composite	CLR	Composite likelihood ratio	The Composite Likelihood Ratio measures an excess of derived alleles across several sites in a single population	Nielsen et al. (2005)
	XP-CLR	Cross population composite likelihood ratio	The Cross Population Composite Likelihood Ratio measures the differences in the excess of derived alleles following population divergence of two populations: the target population and an ancestral population	Chen et al. (2010)
	CMS	Composite of multiple signals	The Composite of Multiple Signals Statistic combines results from F_{ST}, iHS, and XP-EHH to increase power of detecting high frequency, derived alleles between populations	Grossman et al. (2010)
	3P-CLR	3-Population composite likelihood ratio	The 3-population Composite Likelihood Ratio is based on CLR and XP-CLR, except it uses three populations to allow for the detection of selection within each of two populations, or shared selection among the two populations before splitting from a third outgroup population	Racimo (2016)
	SWIF(r)	Sweep inference framework (controlling for correlation)	The Sweep Inference Framework (controlling for correlation) combines a deep learning approach with a composite of the XP-EHH, iHS, and F_{ST} statistics. This method does not require a genome-wide comparison	Sugden et al. (2018)

Allele frequency tests

The joint distribution of allele frequencies in a given population is referred to as the allele frequency spectrum (AFS) or the site frequency spectrum (SFS) (Ewens 1972). The standard neutral model SFS can be predicted theoretically under the assumptions of no selection, no recombination, and that all new genetic variants are unique (Ewens 1972). Selection can act to shift the AFS (Fig. 2) (Fu 1995). Positive selection results in an increase in high frequency alleles and a near absence of intermediate frequency alleles whereas balancing selection results in an excess of alleles at intermediate frequency (Tajima 1989). Demographic events impact the SFS in the same manner as selective events (Marth et al. 2004). Similar to balancing selection, population subdivision reduces the number of highly shared alleles in the population as a whole, thereby increasing the number of intermediate frequency alleles (Tajima 1989). Conversely, population expansion increases the number of singletons and high frequency alleles, much like directional positive selection (Tajima 1989). However, demographic changes are expected to influence all loci across the genome in a similar manner whereas selection tends to be locus-specific (Biswas and Akey 2006; Przeworski et al. 2000). Therefore, to distinguish demography from selection, the SFS at one locus can be compared to the genome-wide average in a given population (Akey et al. 2004; Biswas and Akey 2006).

SFS summary statistics use sequence data to test for departures from neutrality and are useful for identifying loci that have undergone positive or balancing selection (Table 1) (Fay and Wu 2000; Fu and Li 1993; Nielsen et al. 2007; Przeworski et al. 2005; Tajima 1989). Tajima's *D* measures the difference between the number of segregating sites (Θ_W) and the average number of nucleotide differences (π) (Tajima 1989). Under a neutral model, Θ_W and π are equal, and Tajima's *D* is zero. Positive values indicate an excess of intermediate frequency alleles suggesting either balancing selection or population subdivision (Nielsen et al. 2007; Przeworski et al. 2005; Tajima 1989). Negative values indicate an excess of rare alleles and may suggest positive selection or rapid population growth (Nielsen et al. 2007; Przeworski et al. 2005; Tajima 1989). Fu and Li's *D* and *F* statistics compare population mutation rate estimates obtained from derived singleton variants, to Θ_W and π, respectively. Both summary statistics rest on the premise that older variants tend to exist on internal branches, or roots of a phylogenetic tree, while newer variants exist on the external branches (Fu and Li 1993). Therefore, recent positive selection and purifying selection are predicted to result in an excess of variants that lie on the external branches resulting in negative values of the statistics. Balancing selection is predicted to result in positive values of *D* and *F*. One benefit to Fu and Li's *D* is

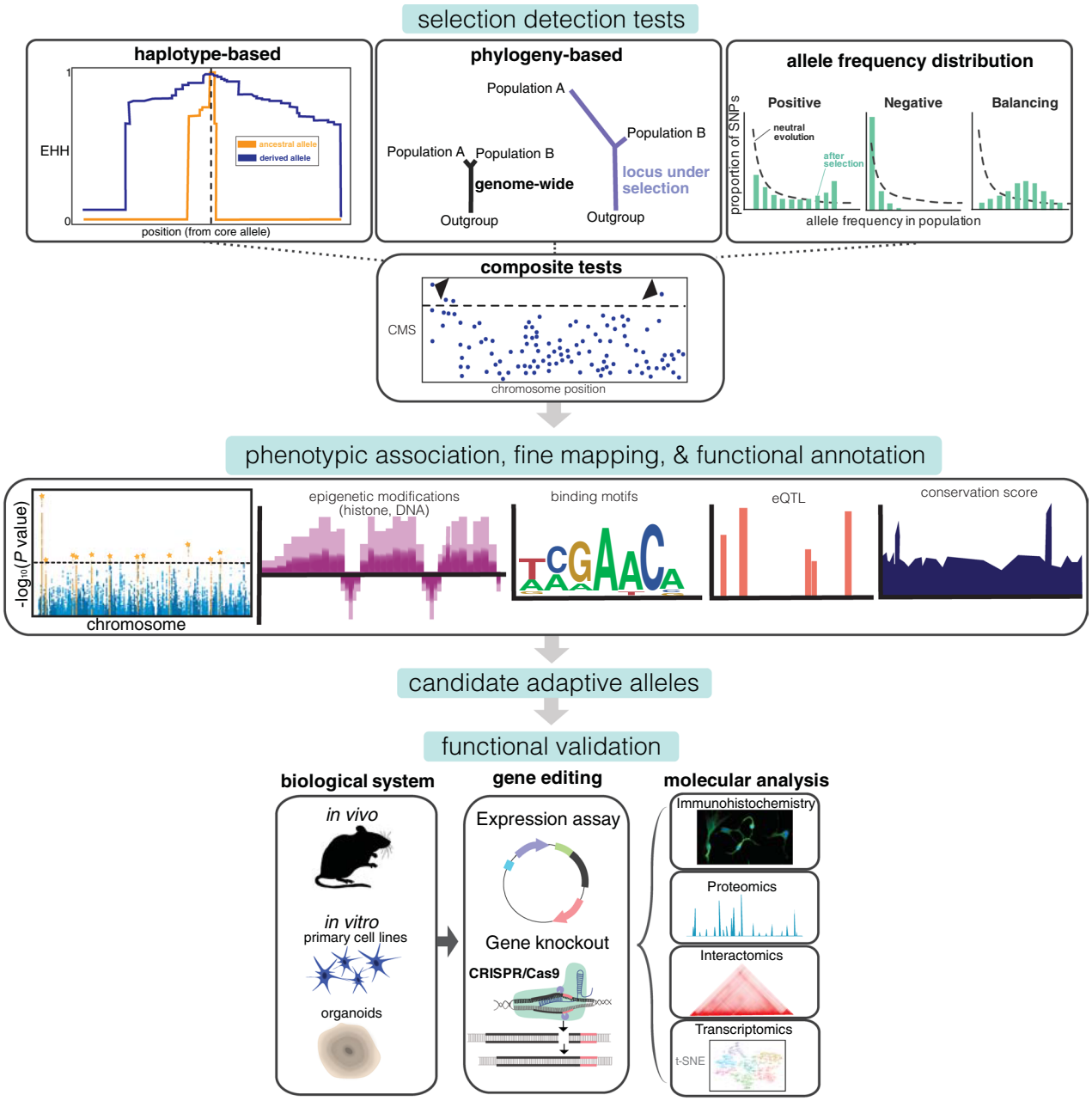

Fig. 2 Reverse genetics selection mapping workflow: First, perform statistical selection tests for a given dataset. Second, integrate with association mapping analyses and fine-mapping approaches to generate a list of high confidence candidate alleles for putative adaptive phenotypes. Third, identify the best research strategy to functionally investigate the candidate variation to model its biological role and inform hypotheses of adaptive human evolution

that it is more sensitive to selective sweeps acting on derived alleles compared to Tajima's *D*. However, selection acting on standing variation can favor ancestral alleles, limiting the ability of these statistics to distinguish positive selection on standing variation. Fay and Wu's *H* statistic measures the amount of high and intermediate frequency derived variants to detect positive selection (Fay and Wu 2000). A negative value of *H* results from an excess of high frequency derived alleles while a positive value implies a lack of intermediate and high frequency derived alleles compared to neutral expectations (Fay and Wu 2000). While SFS summary statistics are popular statistical tests for identifying deviations from neutral evolution within a population, other methods have been developed for comparisons between populations.

Population-specific natural selection can result in allele frequency changes in one population compared to a neutrally evolving population. Wright's fixation index statistic, F_{ST}, measures the amount of genetic differentiation between two or more populations (Table 1) (Wright 1950). Positive selection tends to increase the population differentiation estimates at a selected locus whereas balancing selection tends to decrease F_{ST} (Akey et al. 2002; Andolfatto 2001; Cavalli-Sforza 1966; Wright 1950). Genetic drift also may increase the level of differentiation between populations. However, while positive selection is confined to one or few loci, genetic drift is expected to influence heterogeneity at many unlinked loci across the genome.

Phylogenetic-based tests

Methods that expand on allele frequency differences to explore phylogenetic relationships have been developed to identify population-specific selective events (Table 1, Fig. 2). The locus-specific branch length (LSBL) approach incorporates F_{ST} distances to analyze population-specific allele frequency shifts (Shriver et al. 2004). This method compares F_{ST} between three closely related populations in order to express distances between populations in terms of branch lengths (Shriver et al. 2004). Likewise, the population branch statistic (PBS) uses phylogenetic theory to evaluate changes in allele frequency since population divergence (Yi et al. 2010). Both LSBL and PBS test the null hypothesis that two closely related groups, A and B, have similar branch lengths when compared to an outlier group C. If population A experienced positive selection at a locus, the effect will be a longer branch length than expected. This tree structure will look very different from that of the tree using genome-averaged allelic differentiation values for the three populations.

Composite tests

Composite tests for selection combine a subset of the independent tests for selection described above to increase resolution of the selective signature. Composite methods thus decrease the probability that the signal detected is a false positive and are an effective way to fine-map causal variants. These methods include the composite likelihood ratio (CLR), cross population composite likelihood ratio (XP-CLR), 3-population composite likelihood ratio (3P-CLR), the composite of multiple signals (CMS), and the SWeep Inference Framework (controlling for correlation) (SWIF(r)) (Table 1). CLR identifies an excess of derived alleles across several sites in a single population (Kim and Stephan 2002; Nielsen et al. 2005) whereas XP-CLR does so in two populations (Chen et al. 2010; Vitti et al. 2013). By integrating multiple populations, XP-CLR models allele frequency differences across a chromosome as predicted by the genetic distance from the selected allele (Chen et al. 2010). XP-CLR has greater power to detect selection from standing variation as well as 'ancient' selective sweeps—sweeps that arose following the divergence of two populations, yet ended several thousand generations before the present (Chen et al. 2010). 3P-CLR is a modification of CLR used to distinguish selection events following the divergence of two populations, as well as selection sweeps shared between the two populations prior to the divergence from a third outgroup population (Racimo 2016). The CMS method combines F_{ST}, iHS, and XP-EHH for derived, high frequency alleles in one population compared to other populations (Grossman et al. 2010). This method has been shown to significantly reduce the number of genomic candidates of recent selection for local adaptation in human populations (Grossman et al. 2013, 2010). Similar to CMS, SWIF(r) combines scores from F_{ST}, iHS, and XP-EHH (Sugden et al. 2018). However, this method also incorporates a machine learning mechanism whereby SWIF(r) is trained on simulations from user-selected demographic models (Sugden et al. 2018). By doing so, the probabilistic statistic does not require comparison with a genome-wide distribution, as is the case for CMS, and potentially increases power to localize candidate adaptive loci in both soft and hard selective sweeps (Sugden et al. 2018).

Polygenic adaptation

Polygenic adaptation occurs when multiple adaptive variants across many loci are favored by selection. Selection for polygenic traits only produces slight changes in frequencies of each of the causal alleles (Fig. 1a) (Pritchard et al. 2010). As a consequence, the response of a single allele will be statistically negligible and the signal of selection only can be identified in the cumulative frequency shift of all alleles associated with the phenotype in the population. In other words, polygenic selection for adaptive traits will dampen the signature of selection at each independent locus (Berg and Coop 2014; Haasl and Payseur 2016; Latta 1998), reducing the power of detection by summary statistics (Latta 1998; Le Corre and Kremer 2003). One strategy to increase sensitivity to detection of selection on multi-allelic traits is to apply regression analysis to extended haplotype homozygosity measures (Wiener and Pong-Wong 2011). The regression approach has been proven to be most effective at detecting polygenic selection when allelic diversity is high (Wiener and Pong-Wong 2011).

Limitations for selection mapping on ancient DNA

Selection mapping also has been applied in the burgeoning field of archaeogenetics that seeks to use ancient DNA (aDNA) to resolve questions of hominin evolution, including

the timing of selection events (Knapp and Hofreiter 2010). However, while some have performed aDNA selection scans (Fehren-Schmitz and Georges 2016; Lewandowska et al. 2018; Lindo et al. 2018; Mathieson et al. 2015), applying a selection mapping approach with aDNA sequences is often very difficult given the highly degraded, short-fragment nature of recovered samples (Knapp and Hofreiter 2010). This results in low sequencing coverage, often with only one mapped read at each nucleotide site (Gunther and Nettelblad 2019; Haak et al. 2015; Monroy Kuhn et al. 2018). Given this low sequencing coverage, diploid calls cannot be made, resulting in sequences that cannot be confidently phased and thus are unsuitable for haplotype-based selection tests (Gunther and Nettelblad 2019; Haak et al. 2015; Monroy Kuhn et al. 2018). Therefore, aDNA selection scans are limited to allele frequency-based tests (Gunther and Nettelblad 2019; Haak et al. 2015; Monroy Kuhn et al. 2018). Technological advances in aDNA recovery that improve sample quality and yield upon collection and extraction, as well as methodological improvements of imputation and phasing, may offer more opportunities in the future for selection mapping of aDNA to inform evolutionary models of the chronology and tempo of selective events.

Identifying candidate adaptive variation in human populations

The suite of aforementioned statistical methods provides a means for detecting genomic loci with evidence of positive or balancing selection that can be further dissected to assess their contribution to biological adaptations. Study designs utilizing selection detection methods to hunt down functional adaptive alleles in the genome, and to inform evolutionary hypotheses of human adaptation, can be thought of in two categories: forward genetics or reverse genetics. Those pursuing forward genetics approaches start with a putative adaptive phenotype and uncover the genetic basis (Bigham and Lee 2014; Vitti et al. 2013). In this way, forward genetics seeks to test a specific hypothesis. A classic example of hypothesis-driven, forward genetics using selection mapping to understand human adaptation is the discovery of a causal variant in *HBB* underlying sickle cell anemia that was subsequently found to be under balancing selection in populations living in malaria-endemic regions (Allison 1954; Vitti et al. 2013).

Reverse genetics, on the other hand, seeks to understand the biological function of observed genetic variation, and therefore is hypothesis-generating (Fig. 2) (Hardy et al. 2010). With advancements in genotyping arrays and genome sequencing technology, there have been a burst of whole-genome scans for signals of adaptive evolution in human populations (Crawford et al. 2017b; Eaaswarkhanth et al. 2020; Grossman et al. 2013; Ilardo et al. 2018; Karlsson et al. 2013). Often, these scans identify several candidate genes and/or regulatory regions, or detect signals spanning large segments of the genome (Vitti et al. 2013). In order to pinpoint the causal variant(s) that underlies a putative adaptive phenotype and understand its function, work must first link genetic variation to biological phenotypes.

Identifying putative adaptive phenotypes associated with selection signatures

A common strategy for linking genetic variation to phenotypic variation is to perform genotype–phenotype association analyses. Association analysis can be targeted, such as candidate gene association studies, or genome-wide, such as genome-wide association studies (GWAS) (Rodriguez-Murillo and Greenberg 2008). Population-based association mapping is the process of linking genotypes to phenotypes within a population of unrelated individuals by leveraging LD (Singh and Singh 2015). Study designs include case–control association using a logistic regression and linear regression for quantitative traits (Tak and Farnham 2015). These approaches can help detect genomic regions harboring functional variants, but need to be followed up with further statistical and functional testing to identify the causal variants driving the association (Schaid et al. 2018).

The power of association mapping is dependent upon genotyped markers that lie in regions of strong, detectable LD in the population of interest (Risch and Merikangas 1996; Tak and Farnham 2015). Sample size, effect size, number of causal variants, and allelic heterogeneity all influence the statistical power of detecting an association (Schaid et al. 2018). A major issue with association mapping is that neutral variants linked to the causal variant can be identified as false positives (Tak and Farnham 2015). Also, association mapping is less powered to detect rare functional variants and those of small effect size (Xu et al. 2017). Better phenotypic measurements can sometimes improve the genetic effect size as can increased SNP density and sample size (Schaid et al. 2018). Furthermore, when phenotypic variation correlates with genetic relatedness, or when the study sample has unidentified population structure, markers may produce a false signal of association (Kang et al. 2008). Study designs can control for population structure to reduce the number of false associations (Kang et al. 2008). Given that association mapping can become cumbersome when several causal alleles exist or the relevant phenotypic variation is unavailable, a more efficient strategy to find and prioritize candidate alleles may be preferred. For example, some association methods combine expression data with genetic and phenotypic variables, such as with expression Quantitative Trait Locus (eQTL) mapping (Hormozdiari et al. 2016) and Backward Three-way Association Mapping (BTAM) (Lee et al. 2017).

Integrating selection mapping with association analysis can be a powerful tool to refine loci and localize candidate adaptive alleles (Guo et al. 2018; Johnsson 2018). For example, testing genetic variation in the candidate pigmentation gene, *SLC24A5,* that is under positive selection revealed that the rs1426654-A allele was able to explain ~ 22–32% of total skin color variation in a cohort of individuals from across South Asia (Basu Mallick et al. 2013). Furthermore, selection mapping can be performed before or after genotype–phenotype association analysis, or the two can be statistically combined. However, most combined selection-association mapping studies identify several candidate genes or larger candidate genomic regions. For instance, studies in Bangladeshi individuals from the Ganges River Delta show evidence that cholera has exerted strong selective pressures at ~ 305 candidate genes (Chowdhury et al. 2011; Faruque et al. 2003; Karlsson et al. 2013). Likewise, for putative polygenic adaptations, extensive lists of candidate loci driving overlapping selection and association signals have been compiled, such as for high animal-fat diet adaptation in indigenous Central Siberian populations (lipid and fatty acid metabolism genes *FADS1, FADS2, HADHA, HADHB, MAN1B1, PLD2,* and *AGPAT1*) (Hsieh et al. 2017), or pygmy stature adaptation in African rain forest hunter-gatherers (~ 60 Mb region comprising genes *DOCK3*, *MAPKAP3*, and *CISH)* (Jarvis et al. 2012). While these data represent key initial steps to understanding human genetic adaptation, they nevertheless substantiate the need for further strategies to map causal alleles. These datasets thus serve as rich sources for the fine-mapping and prioritization of candidate adaptive variants for functional testing.

Fine mapping and functional annotation to prioritize candidate adaptive alleles

'Fine mapping' is a strategy to localize causal variants from association and/or selection mapping hits through statistical, bioinformatic, and/or functional approaches (Schaid et al. 2018). Common statistical fine mapping methods, developed for GWAS hits, include heuristic methods (analyze pairwise correlations of each surrounding SNPs with the identified lead SNP), penalized regression models (use cross-validation to randomly partition into training sets to select SNPs to use in the regression), and Bayesian methods (probabilistic modeling) (Schaid et al. 2018). Fine mapping methods include CMS (described above), 'selection detection by conditional coalescent tree' (SCCT) (Wang et al. 2014), 'integrated selection of allele favored by evolution' (iSAFE, exploits the shoulders of sweeps with high performance up to 5 Mbp regions) (Akbari et al. 2018), and Fine-Mapping of Adaptive Variation (FineMAV, combines population differentiation, derived allele frequency, and molecular functional annotation) (Szpak et al. 2018).

Functional annotation of variants present within candidate genes or non-coding regulatory regions can help rank and prioritize variants for functional investigation. It can be used to inform statistical fine mapping or be assessed following fine mapping analyses (Schaid et al. 2018). Annotation criteria include mutation type (substitution, insertion, deletion, duplication, inversion, translocation), effect on coding sequence (missense, nonsense, frameshift, splice), and non-coding regulatory element type (promoter, enhancer, silencer, insulator) (Ip et al. 2019; Zhou and Zhao 2018). Variants also can be prioritized based on where they localize in the functional element, such as within a particular protein domain, splicing site, transcription factor binding site, among others (Ip et al. 2019; Zhou and Zhao 2018).

Functional annotations can be made from searching published literature, mining publicly available functional annotation databases, and using free bioinformatics tools. CAUSALdb is a free database that offers uniform fine-mapping of GWAS summary data pulled from 3000 studies using three state-of-the-art tools (including PAINTOR, CAVIARBF and FINEMAP), combined with functional annotation, offering users a browser view of variant-, gene-, and trait-level causal relations (mulinlab.tmu.edu.cn/causaldb, (Wang et al. 2020). Some tools have been curated for improving variant prioritization based on annotation data and functional prediction, such as Variant Prioritization Ordering Tool (VPOT) (Ip et al. 2019). The Encyclopedia of DNA Elements (ENCODE) Project portal (encodeproject.org) is a great resource for genome-wide annotations of functional elements, including a web interface called SCREEN, or Search for Candidate cis-Regulatory Elements by ENCODE (screen.wenglab.org). The UCSC Genome Browser is another excellent resource to simultaneously visualize functional annotations across a genomic interval from several databases, with annotation tracks for cross-species conservation scores, structural variants, phenotype-associated variants from the Online Mendelian Inheritance in Man (OMIM, omim.org), ClinVar (ncbi.nlm.nih.gov/clinvar/), UniProt (uniprot.org), and GWAS studies, to name a few. The browser also includes tracks for common SNPs from catalogs such as dbSNP (ncbi.nlm.nih.gov/projects/SNP) and the Genome Aggregation Database (GnomAD) (gnomad.broadinstitute.org). The GnomAD browser, developed by the BROAD Institute, compiles whole genome and exome sequencing data from disease-specific and population genetic cohorts, totaling 213,158 unrelated individuals, offering variant-level annotations (gnomad.broadinstitute.org). The Genotype-Tissue Expression (GTEx) Portal (gtexportal.org) provides extensive information on the relationship between genetic variants and human gene regulation/expression by using global RNA expression data across several tissue types and eQTL analysis. The WashU Epigenome Browser (epigenomegateway.wustl.edu) is a great

visualization tool of high-throughput epigenomic and expression data from diverse adult and embryonic cell types, including annotated regions of open chromatin from DNAse hypersensitivity assays as well as histone modifications associated with gene repression or activation.

Insight into the molecular mechanism of an adaptive change can be gained from determining where candidate variants localize with respect to protein structure and protein sequence. Some popular biomolecular structure databases include Protein Data Bank (PDB, rcsb.org/pdb), Structure Classification of Proteins (SCOP, scop.berkeley.edu), and 'Class, Architecture, Topology' (CATH, cathdb.info). Three-dimensional structural analysis can be helpful to model the variant effect on protein structure, which can be achieved using free bioinformatics tools, such as PyMol (pymol.org/2/) and Missense3D (sbg.bio.ic.ac.uk/~missense3d/). PDB offers a suite of other available molecular graphics software (rcsb.org/pages/thirdparty/molecular_graphics). Useful resources to analyze protein function and domain information include PFam (pfam.xfam.org), Prosite (prosite.expasy.org), and PRINTS (130.88.97.239/PRINTS/index.php). Lastly, phenotype-drive mutagenesis screens, using different genome editing tools, are a rich source of annotated variants with visible phenotypes. Several high-throughput mutagenesis screens have been performed in diverse human cells, and available databases include GenomeRNAi (genomernai.org) and GenomeCRISPR (omictools.com/genomecrispr-tool). Functional annotation and bioinformatic modeling of genetic variants across diverse criteria using different online resources is a cost- and time-effective way to prioritize candidate variants for additional allele frequency tests and for functional investigation. Together these data can build a case for putative adaptive function in human evolution.

Major challenges remain for fine-mapping and variant prioritization despite the growing number of functional annotation resources and tools. This is particularly the case when results from integrated selection and association mapping studies detect extensive and/or numerous genomic loci; or when several causal variants of small effect contribute to the statistical signature of selection (Akbari et al. 2018; Schaid et al. 2018; Szpak et al. 2018; van de Bunt et al. 2015). While it remains an ongoing challenge to pinpoint the functional allele(s) driving the selection signal, we discuss successful examples in human populations below including infectious disease resistance and/or susceptibility, climatic adaptations, and dietary adaptations (Fig. 3, Table 2).

Candidate adaptive alleles identified in human populations

Infectious disease adaptations: malaria Searching for signatures of selection to identify susceptibility alleles for infectious disease is an appealing strategy as pathogens are hypothesized to have exerted strong selective pressures on evolving modern humans (Fumagalli et al. 2011; Haldane 1932; Karlsson et al. 2014). Many genes subject to local positive natural selection (e.g. *DARC*) or balancing natural selection (e.g. *G6PD, SLC4A1*) are associated with susceptibility to infectious disease (Bamshad et al. 2002; Hughes and Yeager 1998; Sabeti et al. 2002a). Malaria, perhaps the most extensively studied human infectious disease, imparts one of the strongest selective pressures on modern humans (Evans and Wellems 2002). Co-evolution with the malaria parasite has led homozygous deleterious alleles that are associated with particular blood disorders to confer a selective advantage in the heterozygous state (Kwiatkowski 2005; Luzzatto 2012). The most well-known adaptive allele that confers resistance to *Plasmodium falciparum* malaria is the sickle cell allele (HbS) of the beta-globin locus, *HBB*, wherein individuals heterozygous for the HbS allele are protected from malaria invasion due to the malformation of a portion of the erythrocytes, yet do not present with severe sickle cell anemia (Allison 1954; Haldane 1949; Piel et al. 2010). The HbS allele of SNP rs334, caused by a missense variant c.20A>T(p.Glu7Val) in *HBB*, may function to reduce malarial virulence by reducing surface expression of *P. falciparum* erythrocyte membrane protein-1 (PfEMP-1) on infected erythrocyte cells, thereby impairing their cytoadherence to microvascular endothelial cells, a property necessary for the parasite to escape bloodstream removal by the spleen (Cholera et al. 2008).

In addition to the well-known *HBB* example, additional candidate alleles under balancing selection have been identified that underlie malarial resistance adaptations. Hypomorphic alleles at the glucose-6-phosphate dehydrogenase locus (*G6PD)* on the X chromosome that cause G6PD enzyme deficiency are the genetic basis for a number of hemopathologies in humans (Gregg and Prchal 2018). Among these alleles, the rs137852331-G (or "*G6PD A*") allele and the rs1050828-A (or "*G6PD A-")* allele are under balancing selection for partial malaria resistance in sub-Saharan African populations, while the rs5030868-T (or "*G6PD Mediterranean*") allele may be under balancing selection in Southern European, Middle Eastern, and Indian populations (Kwiatkowski 2005; Saunders et al. 2002, 2005; Tishkoff et al. 2001; Verrelli et al. 2002). The *G6PD A-* allele results in the missense change c.292G > A(p.Val98Met) that reduces G6PD activity by 82% compared to the ancestral allele and is associated with a 46–58% reduction in severe malaria risk in both heterozygous females and hemizygous males (Ruwende et al. 1995). Southeast Asian Ovalocytosis (SAO) is another blood disease resulting in oval-shaped erythrocytes that is protective against severe vivax malaria (Jarolim et al. 1991). A 27 base-pair deletion (Δ27) in the *SLC4A1* gene which encodes an anion transport protein is responsible for this phenotypic trait, and while the *SLC4A1Δ27* allele

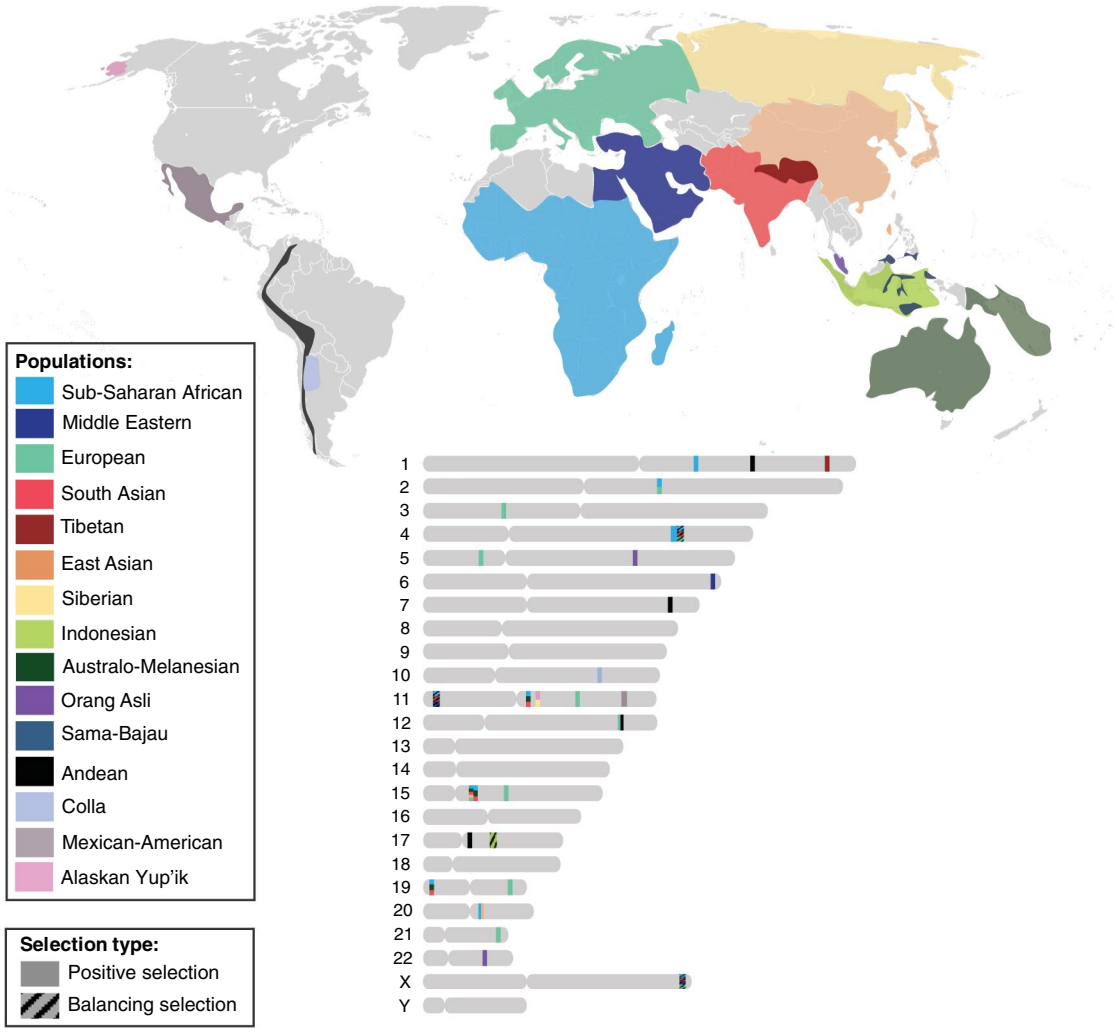

Fig. 3 Map illustrating the suite of discovered candidate adaptive alleles within diverse human populations highlighted in this review. These alleles are represented by rectangular bands, color-coordinated by the general geographic region of the study population(s), and mapped to the respective broader chromosomal region. Selection type is indicated by pattern: solid (positive selection) and striped (balancing selection). Multi-colored bands indicate loci where adaptive alleles are found in multiple populations. Note: this is not a comprehensive map of all identified candidate adaptive alleles in humans

is protective against malaria and typically asymptomatic in heterozygous individuals, it is lethal in its homozygous state (Jarolim et al. 1991; Paquette et al. 2015; Wilder et al. 2009). Indonesian populations exhibit an excess of high-frequency, derived variants around *SLC4A1Δ27* (Wilder et al. 2009), suggesting that natural selection has acted recently on this locus.

The Duffy antigen receptor for chemokine, *DARC* (or *ACKR1*), is another target of selection from malarial pressure (Oliveira et al. 2012). *DARC* encodes the transmembrane glycoprotein Duffy antigen (*Fy*) (Hadley and Peiper 1997), which exhibits allelic differentiation by geographic region with near fixation for the Duffy-null allele, *Fy-*, in the majority of sub-Saharan Africa (Livingstone 1984). The Duffy-null allele is due to a A>G substitution at rs2814778 in the GATA promoter of *DARC*, and individuals homozygous for Duffy-null (*Fy-/Fy-*) are resistant to erythrocyte invasion by *Plasmodium vivax* malaria (Miller et al. 1976). The selection signature around *DARC* extends ~ 6 kb upstream of the 5′ UTR (Hamblin and Di Rienzo 2000). Levels of sequence diversity and site frequency spectrum results indicate directional positive selection at *DARC* in sub-Saharan African populations (the Mandinka, Beti, Hausa, Mbuti, and Luo) when compared to sequence data from non-African groups (Italian, Han Chinese, Pakistani) that do not fit simple models of selection (Hamblin and Di Rienzo 2000; Hamblin et al. 2002). These results support the hypothesis that the fixation of the Duffy-null allele in

Table 2 Candidate human adaptive alleles discussed in this review and illustrated in Fig. 3

Category	Variant Type	Candidate Adaptive Allele	Functional Annotation	Chromosomal position (Hg38)	Gene (within or nearby)	Population(s)	Selection Mode	Trait	References
Immunity	Substitution	rs2814778-G	5′ UTR[1]	Chr1:159204893	*DARC*	Sub-Saharan African, Malagasy	Positive	Infectious disease resistance (Malaria)	Hamblin and Di Rienzo (2000), Hodgson et al. (2014)
Climatic	Substitution	rs11578671-A	Intergenic	Chr1:189811662	*BRINP3*	High-Altitude Andean	Positive	Polycythemia tolerance from high altitude adaptation	Crawford et al. (2017a)
Climatic	Substitution	rs12097901-C	Exonic: missense (p.Cys127Ser)	Chr1:231421509	*EGLN1*	High-Altitude Tibetan	Positive	High altitude adaptation	Lorenzo et al. (2014)
Climatic	Substitution	rs186996510-G	Exonic: missense (p.Asp4Glu)	Chr1:231421877	*EGLN1*	High-Altitude Tibetan	Positive	High altitude adaptation	Lorenzo et al. (2014)
Dietary	Substitution	rs41525747-G	Intronic	Chr2:135851073	*MCM6*	African Pastoralist (Tanzania, Kenya, Sudan)	Positive	Lactase persistance	Tishkoff et al. (2007)
Dietary	Substitution	rs4988235-T	Intronic	Chr2:135851076	*MCM6*	Northern European	Positive	Lactase persistance	Enattah et al. (2002)
Dietary	Substitution	rs41380347-G	Intronic	Chr2:135851081	*MCM6*	African Pastoralist (Tanzania, Kenya, Sudan)	Positive	Lactase Persistance	Tishkoff et al. (2007)
Dietary	Substitution	rs145946881-C	Intronic	Chr2:135851176	*MCM6*	African Pastoralist (Tanzania, Kenya, Sudan)	Positive	Lactase persistance	Tishkoff et al. (2007)
Dietary	Substitution	rs182549-A	Intronic	Chr2:135859184	*MCM6*	Northern European	Positive	Lactase persistance	Enattah et al. (2002)
Immunity	Deletion	*CCR5Δ32* (rs333)	Exonic: frameshift deletion	Chr3:46373453	*CCR5*	European	Positive	Infectious disease resistance (HIV)	Liu et al. (1996), Samson et al. (1996)
Immunity	CNV[2]	DUP4	Fusion protein: GYPB/GYPA	Chr4:143870000–144150000	Affects glycoporin locus	East African	Positive	Infectious disease resistance (Malaria)	Algady et al. (2018)
Immunity	Substitution	rs7687256-T	Exonic: missense (p.Glu24Gly)	Chr4:144120555	*GYPA*	Sub-Saharan African, South Asian, European	Balancing	Infectious disease resistance (Malaria)	Bigham et al. (2018), Ko et al. (2011)
Immunity	Substitution	rs7682260-A	Exonic: missense (p.Ser20Leu)	Chr4:144120567	*GYPA*	Sub-Saharan African, South Asian, European	Balancing	Infectious disease resistance (Malaria)	Bigham et al. (2018), Ko et al. (2011)
Immunity	Substitution	rs40401-T	Exonic: missense (p.Pro27Ser)	Chr5:132060785	*IL3*	Orang Asli, Ghanaian, Nigerian, African American	Positive	Infectious disease resistance (Malaria)	Liu et al. (2015)
Immunity	Substitution	rs2243250-T	Intergenic	Chr5:132673462	*IL4*	Orang Asli	Positive	Infectious disease resistance (Malaria)	Liu et al. (2015)

Table 2 (continued)

Category	Variant Type	Candidate Adaptive Allele	Functional Annotation	Chromosomal position (Hg38)	Gene (within or nearby)	Population(s)	Selection Mode	Trait	References
Immunity	Substitution	rs2070874-T	5′ UTR	Chr5:132674018	*IL4*	Orang Asli	Positive	Infectious disease resistance (Malaria)	Liu et al. (2015)
Climatic	Substitution	rs16891982-G	Exonic: missense (p.Phe374Leu)	Chr5:33951588	*MATP*	European	Positive	Skin pigmentation adaptation	Lao et al. (2007), Norton et al. (2007)
Climatic	Substitution	rs3008052-T	Intronic	Chr6:165653494	*PDE10A*	Sama-Bajau	Positive	Deep-sea diving adaptation	Ilardo et al. (2018)
Dietary	Substitution	rs77768615-C	Intronic	Chr7:141999184	*MGAM*	Andean	Positive	High starch diet adaptation	Lindo et al. (2018)
Dietary	Substitution	rs7085104-G	Intronic	Chr10:102869116	*AS3MT*	Colla	Positive	Arsenic water adaptation	Eichstaedt et al. (2015)
Dietary	Substitution	rs1046778-C	Intronic	Chr10:102901727	*AS3MT*	Colla	Positive	Arsenic water adaptation	Eichstaedt et al. (2015)
Metabolic	Substitution	rs3135506-A	Exonic: missense (p.Ser19Trp)	Chr11:116791691	*APOA5*	Mexican American	Positive	Increased dyslipidemia risk (Gene-environment mismatch)	Ko et al. (2014)
Metabolic	Substitution	rs662799-C	Intergenic	Chr11:116792991	*APOA5*	Mexican American	Positive	Increased dyslipidemia risk (Gene-environment mismatch)	Ko et al. (2014)
Metabolic	Substitution	rs662799-C	Intergenic	Chr11:116792991	nearby *LPL*	Mexican American	Positive	Increased dyslipidemia risk (Gene-environment mismatch)	Ko et al. (2014)
Metabolic	Substitution	rs139961185-A	Intronic	Chr11:116936627	*SIK3*	Mexican American	Positive	Increased dyslipidemia risk (Gene-environment mismatch)	Ko et al. (2014)
Immunity	Substitution	rs334-T	Exonic: missense (p.Glu7Val)	Chr11:5227002	*HBB*	Sub-Saharan African, Middle Eastern, South Asian	Balancing	Infectious disease resistance (Malaria)	Allison (1954); Piel et al. (2010)
Climatic	Substitution	rs11230664-C	Intronic	Chr11:61308900	*DDB1*	African, South Asian, Australo-Melanesian	Positive	Skin pigmentation adaptation	Crawford et al. (2017b)
Climatic	Substitution	rs7948623-T	3′ UTR	Chr11:61369675	*TMEM138*	African, South Asian, Australo-Melanesian	Positive	Skin pigmentation adaptation	Crawford et al. (2017b)

Table 2 (continued)

Category	Variant Type	Candidate Adaptive Allele	Functional Annotation	Chromosomal position (Hg38)	Gene (within or nearby)	Population(s)	Selection Mode	Trait	References
Climatic	Substitution	rs80356779-T	Exonic: missense (p.Pro479Leu)	Chr11:68780662	*CPT1A*	Siberian	Positive	Cold climate adaptation	Clemente et al. (2014), Cardona et al. (2014)
Climatic	Substitution	rs80356779-T	Exonic: missense (p.Pro479Leu)	Chr11:68780662	*CPT1A*	Alaskan Yup'ik	Positive	Cold climate adaptation	Clemente et al. (2014)
Climatic	Substitution	rs1042602-A	Exonic: missense (p.Ser192Tyr)	Chr11:89178528	*TYR*	European	Positive	Skin pigmentation adaptation	Norton et al. (2007)
Immunity	Substitution	rs34137742-T	Intronic	Chr12:112910856	*OAS1*	European	Positive	Infectious disease susceptibility (West Nile Virus)	Bigham et al. (2011)
Climatic	Substitution	rs10744822-C	Intronic	Chr12:114346235	*TBX5*	High-Altitude Andean	Positive	Polycythemia tolerance from high altitude adaptation	Crawford et al. (2017a)
Climatic	Substitution	rs1800414-G	Exonic: missense (p.His615Arg)	Chr15:27951891	*OCA2*	East Asian	Positive	Skin pigmentation adaptation	Lao et al. (2007)
Climatic	Substitution	rs1800404-C	Exonic: synonymous	Chr15:27990627	*OCA2*	African, South Asian, Australo-Melanesian	Positive	Skin pigmentation adaptation	Crawford et al. (2017b)
Climatic	Substitution	rs1448484-T	Intronic	Chr15:28038295	*OCA2*	European, East Asian	Positive	Skin pigmentation adaptation	Lao et al. (2007)
Climatic	Substitution	rs6497271-A	Intronic	Chr15:28120285	*HERC2*	African, South Asian, Australo-Melanesian	Positive	Skin pigmentation adaptation	Crawford et al. (2017b)
Climatic	Substitution	rs4932620-T	Intronic	Chr15:28269135	*HERC2*	African, South Asian, Australo-Melanesian	Positive	Skin pigmentation adaptation	Crawford et al. (2017b)
Climatic	Substitution	rs1426654-A	Exonic: missense (p.Thr111Ala)	Chr15:48134287	*SLC24A5*	European	Positive	Skin pigmentation adaptation	Lamason et al. (2005), Norton et al. (2007)
Climatic	Substitution	rs34913965-T	Intronic	Chr17:27806151	*NOS2*	High-Altitude Andean	Positive	Polycythemia tolerance from high altitude adaptation	Crawford et al. (2017a)
Immunity	Deletion	*SLC4A1Δ27* (rs769664228)	Exonic: in-frame deletion	Chr17:44258043	*SLC4A1*	Indonesian	Balancing	Infectious disease resistance (Malaria)	Jarolim et al. (1991), Paquette et al. (2015), Wilder et al. (2009)

Table 2 (continued)

Category	Variant Type	Candidate Adaptive Allele	Functional Annotation	Chromosomal position (Hg38)	Gene (within or nearby)	Population(s)	Selection Mode	Trait	References
Climatic	Substitution	rs56203814-T	Exonic: synonymous	Chr19:3544894	*MFSD12*	African, South Asian, Australo-Melanesian	Positive	Skin pigmentation adaptation	Crawford et al. (2017b)
Climatic	Substitution	rs10424065-T	Intronic	Chr19:3545024	*MFSD12*	African, South Asian, Australo-Melanesian	Positive	Skin pigmentation adaptation	Crawford et al. (2017b)
Climatic	Substitution	rs6510760-A	Intergenic, regulatory	Chr19:3565255	*MFSD12*	African, South Asian, Australo-Melanesian	Positive	Skin pigmentation adaptation	Crawford et al. (2017b)
Climatic	Substitution	rs112332856-C	Intergenic, regulatory	Chr19:3565601	*MFSD12*	African, South Asian, Australo-Melanesian	Positive	Skin pigmentation adaptation	Crawford et al. (2017b)
Immunity	Substitution	rs2304207-C	Intronic	Chr19:49664469	*IRF3*	European	Positive	Infectious disease susceptibility (West Nile Virus)	Bigham et al. (2011)
Climatic	Substitution	rs6058017-G	Intronic	Chr20:34269192	*ASIP*	African	Positive	Skin pigmentation adaptation	Norton et al. (2007)
Climatic	Substitution	rs4911178-A	Intronic, Regulatory	Chr20:35364817	*GROW1* (*GDF5* enhancer)	East Asian	Positive	Body size and proportion	Capellini et al. (2017)
Immunity	Substitution	rs7280422-G	Intronic	Chr21:41429640	*MX1*	European	Positive	Infectious disease susceptibility (West Nile Virus)	Bigham et al. (2011)
Immunity	Substitution	rs2071748-A	Intronic	Chr22:35381625	*HMOX1*	Orang Asli	Positive	Infectious disease resistance (Malaria)	Liu et al. (2015)
Immunity	Substitution	rs8139532-G	Intronic	Chr22:35383575	*HMOX1*	Orang Asli	Positive	Infectious disease resistance (Malaria)	Liu et al. (2015)
Immunity	Substitution	rs5030868-T	Exonic: missense (p.Ser188Phe)	ChrX: 154534419	*G6PD*	Southern European, Middle Eastern, Indian	Balancing	Infectious disease resistance (Malaria)	Tishkoff et al. (2001), Verrelli et al. (2002)
Immunity	Substitution	rs1050828-A	Exonic: missense (p.Val98Met)	ChrX: 154536002	*G6PD*	Sub-Saharan African	Balancing	Infectious disease resistance (Malaria)	Saunders et al. (2002, 2005), Tishkoff et al. (2001), Verrelli et al. (2002)

Table 2 (continued)

Category	Variant Type	Candidate Adaptive Allele	Functional Annotation	Chromosomal position (Hg38)	Gene (within or nearby)	Population(s)	Selection Mode	Trait	References
Immunity	Substitution	rs137852331-G	Exonic: missense (p.Asn165Asp)	ChrX:154534489	*G6PD*	Sub-Saharan African	Balancing	Infectious disease resistance (Malaria)	Saunders et al. (2002, 2005), Tishkoff et al. (2001), Verrelli et al. (2002)

[1] *UTR* untranslated region

[2] *CNV* copy number variant

mainland sub-Saharan Africa was the result of selective pressures from vivax malaria. Additionally, multiple Malagasy populations in Madagascar show high frequencies of the Duffy-null allele, with evidence for recent positive selection in the Merina Highlanders (Hodgson et al. 2014).

The *GYPA* locus encodes a ligand of the glycophorin A receptor expressed on the cell surface of erythrocytes that *P. falciparum* malaria uses to gain cell entry (Camus and Hadley 1985; Sim et al. 1994; Tolia et al. 2005). From targeted selection tests, *GYPA* shows patterns of variation supporting balancing selection in both malaria-endemic populations as well as in European populations (Bigham et al. 2018; Ko et al. 2011). Candidate adaptive alleles driving balancing selection at *GYPA* are the ancestral alleles rs7682260-G and rs7687256-C, and derived alleles rs7682260-A and rs7687256-T, which encode the M and N antigens on GYPA, respectively (Bigham et al. 2018; Ko et al. 2011). Alleles rs7682260-A and rs7687256-T result in missense changes p.Ser20Leu and p.Glu24Gly, respectively. In addition, the glycophorin locus (spanning genes *GYPE, GYPB,* and *GYPA*) is known to undergone copy number variation and gene conversion in the human population (Algady et al. 2018; Leffler et al. 2017; Suktitipat et al. 2014). A structural variant called DUP4, likely the result of a series of six copy number changes at the glycophorin locus, generates a GYPB-GYPA fusion protein (Leffler et al. 2017). The DUP4 allele is under positive selection in East African populations, likely conferring protection against *P. falciparum* infection by altering the parasite's ability to interact with erythrocyte surface receptors (Algady et al. 2018). Differential signals of positive selection also have been discovered across Orang Asli populations from peninsular Malaysia in the immune response interleukin genes *IL3* and *IL4*, as well as genes *HMOX1*, *LTA*, *TNF*, *CDH13, NOS2*, *IRF1*, and *FAS,* supporting the hypothesis that these groups have undergone convergent adaptation to counteract malaria (Liu et al. 2015). Fine mapping of candidate alleles within *IL3* and *IL4* genes that are under positive selection in the Che Wong population identified top candidate SNPs rs40401, rs2243250, and rs2070874 (Liu et al. 2015). The rs40401-T allele of *IL3*, which causes a missense change c.79C>T(p.Pro27Ser), is observed at high frequency in the Che Wong, as well as in Ghanaian, Nigerian, and African American populations (Liu et al. 2015). Furthermore, epidemiological studies found a 33% protective effect from recurrent malaria episodes in Ghanaians (Meyer et al. 2011). Top candidate alleles of *IL4* (rs2243250-T and rs2070874-T) lie within the promoter, and are associated with higher anti-malarial antibody levels (Liu et al. 2015). The only other candidate alleles identified were in the Jakun population in the heme oxygenase regulator *HMOX1* gene (rs2071748-A and rs8139532-G) (Liu et al. 2015), both were previously found to be associated with cerebral malaria protection (Sambo et al. 2010). Altogether, the

above examples from diverse populations demonstrate that malaria has been a robust evolutionary pressure resulting in localized adaptations in immune response.

Infectious disease adaptations: HIV-1 Selection mapping also can be used to identify risk alleles to modern pathogens that were not the past selective agent (Karlsson et al. 2014). This is particularly relevant given the high level of redundancy of cellular and molecular protective activity within the immune system (Nish and Medzhitov 2011). This evolved redundancy arms the immune system with surplus compensatory mechanisms, for example, when loss of function variants disrupt certain immune response pathways (Fischer and Rausell 2016). In this way, host immune response utilizes existing genomic and molecular variation to respond to novel pathogens. Favored host genetic variation from past selective pressures from ancient pathogens can influence host immune resistance or susceptibility phenotypes in the face of novel pathogens. One well-studied immune locus under natural selection in human populations is the gene *CCR5,* which encodes a protein expressed on the cell surface of white blood cells (Libert et al. 1998; Novembre et al. 2005; Stephens et al. 1998). A 32 base-pair deletion (Δ32) at this locus, *CCR5Δ32*, is at high frequencies and under selection in Europe and western Asia (Novembre et al. 2005). Interestingly, the *CCR5Δ32* allele confers resistance to infection by Human Immunodeficiency Virus-1 (HIV-1) by abrogating expression of the CCR5 chemokine receptor exploited by HIV-1 for entry into CD4 + T cells (Liu et al. 1996; Samson et al. 1996). The HIV-1 lentivirus evolved less than 100 years ago in Africa from simian immunodeficiency virus, and is therefore not the selective pressure that previously acted on the *CCR5* locus (Sharp et al. 2001). Through the reconstruction of historical events, scholars hypothesize a different pathogen, such as smallpox, drove the signature of selection observed at this locus (Galvani and Slatkin 2003). The story of the *CCR5Δ32* allele demonstrates how previously adaptive genetic variation can influence current susceptibility to modern infectious diseases. Further, it substantiates the use of selection mapping for identifying genes and/or regulatory elements that harbor resistance and/or susceptibility alleles to modern pathogens.

Infectious disease adaptations: West Nile Virus West Nile Virus (WNV) is second modern pathogen for which susceptibility alleles have been identified by prioritizing genes with strong signatures of natural selection (Bigham et al. 2011; Lim et al. 2009; Yakub et al. 2005). WNV is responsible for highly variable infection outcomes across populations wherein infected individuals may be asymptomatic, present with West Nile fever, or develop potentially fatal West Nile neuroinvasive disease (Bigham et al. 2011). A case–control association study informed by selection mapping to identify host genetic susceptibility factors to WNV infection among Europeans found strong associations at intronic SNPs in *IRF3* (rs2304207-C allele) and *MX1* (rs7280422-G allele) with symptomatic WNV infection, as well as in *OAS1* (rs34137742-T allele) with severe infection (Bigham et al. 2011). This research illustrates how selection mapping can be used to prioritize candidate genes for case–control association testing to localize susceptibility alleles.

Climatic adaptations: high altitude High altitude, defined as regions lying above 2,500 m, represents one of the most marginalized and harsh environments inhabited by humans (Niermeyer et al. 2001). There are three high-altitude regions where humans have resided for millennia: the Tibetan Plateau, the Andean Altiplano, and the Semien Plateau of Ethiopia (Beall 2013). Populations from these regions display physiological and genetic adaptations to overcome reduced oxygen levels (Bigham and Lee 2014; Scheinfeldt and Tishkoff 2013; Simonson et al. 2012). Genome-wide and candidate gene research has identified compelling evidence for selection on the hypoxia-inducible factor (HIF) pathway and hypoxia-related genes (Alkorta-Aranburu et al. 2012; Beall et al. 2010; Bigham et al. 2010; Scheinfeldt et al. 2012; Simonson et al. 2010; Yi et al. 2010). Candidate alleles in *EGLN1, EPAS1,* and *PRKAA1* are suggested to underlie altitude-adaptive phenotypes including hemoglobin concentration, birth weight, and maximal oxygen consumption during exercise (Beall et al. 2010; Bigham et al. 2010, 2014; Brutsaert et al. 2019; Eichstaedt et al. 2017; Simonson et al. 2010; Yi et al. 2010). For example, high frequency candidate alleles in *EGLN1,* rs12097901-C and rs186996510-G (resulting in missense changes c.380G>C(p.Cys127Ser) and c.12C>G(p.Asp4Glu), respectively), are associated with hemoglobin concentration in Tibetan highlanders (Lorenzo et al. 2014). In addition to high altitude, human adaptation to hypoxic conditions may have occurred in the Bajau deep-sea divers of Southeast Asia. An intronic SNP in *PDE10A* (rs3008052) that is found within a peak positive selection signal in this population, shows significant association with enlarged spleen size (Ilardo et al. 2018). The discovery and functional investigation of adaptive alleles that mitigate hypoxia has translational implications for understanding and treating the disease biology caused by chronic hypoxia, including cases of ischemic heart disease, cerebrovascular disease, anemia, chronic obstructive pulmonary disease, and pulmonary hypertension.

Climatic adaptations: cold climate Another extreme environmental pressure that challenges human survival is cold climate. Siberia is the coldest region inhabited by humans with an average annual temperature roughly 5 °C. Archaeological data estimate first human arrivals into the region around 45–40 kya (Goebel 1999; Goebel et al. 2008).

Although cultural adaptations can help buffer humans from the physiological stressors of extreme cold, biological adaptations, such as increased levels of thermogenic brown adipose tissue (Daanen and Van Marken Lichtenbelt 2016), are necessary to withstand continued sub-zero temperatures. Genome-wide selection scans on indigenous populations from Siberia using iHS, XP-EHH, and PBS discovered a 3 Mb region on chromosome 11, containing ~ 79 genes with strong positive selection signals (Cardona et al. 2014; Clemente et al. 2014). While genes involved in the regulation of energy and metabolism (*CPT1A, LRP5,* and *THADA)* as well as the smooth muscle contraction, *PRKG1,* were identified in this scan, candidate alleles driving this signal we not identified (Cardona et al. 2014). However, further selection mapping work using whole genome sequencing data from Canadian and Greenland Inuit populations identified a nonsynonymous variant [c.1436C>T (p.Pro479Leu)] at the rs80356779 SNP in *CPT1A*, which is essential for fatty acid oxidation (Clemente et al. 2014). Given the rs80356779-T allele also is observed at a very high frequency in the Northeast Siberian populations from Cardona et al., 2014, it is a top candidate causal allele driving the strong signal of positive selection from cold exposure (Clemente et al. 2014). Functional evidence shows that the *CPT1A* c.1436C>T (p.Pro479Leu) missense variant decreases fatty-acid oxidation and ketogenesis, decreases mitochondrial inhibition of malonyl-CoA on fatty acid β-oxidation, and is associated with increased high-density lipoprotein cholesterol with reduced adiposity in the Alaskan Yup'iks population (Clemente et al. 2014; Greenberg et al. 2009; Lemas et al. 2012). Together these data suggest an adaptive metabolic role based on gene-environment interaction (Clemente et al. 2014; Collins et al. 2010). Functional models seeking to recapitulate an adaptive function of this allele based on different environmental parameters of temperature and diet will be important to shed light on gene-environment interaction hypotheses. In addition, further fine mapping and functional investigation of other candidate adaptive alleles underlying cold adaptation in humans may help to uncover regulatory mechanisms of human adipogenesis, with potential to inform medical research of certain metabolic diseases, such as obesity.

In addition to metabolic adaptations, it is widely hypothesized that human body size and body proportions have adapted to temperature exposure for proper thermoregulation. Signatures of positive selection surrounding height loci have been identified (Amato et al. 2011; Voight et al. 2006; Wu et al. 2012). One interesting selection candidate for short stature is *GDF5,* which encodes the growth differentiation factor 5 protein involved in skeletal, joint, and central nervous system development (Buxton et al. 2001; Francis-West et al. 1999). Among East Asian populations, *GDF5* shows evidence of positive selection and variation in this gene is associated with both decreased stature and osteoarthritis risk (Wu et al. 2012). Using a CMS approach, a strong selection peak was observed surrounding the 3′ UTR of *GDF5* (Capellini et al. 2017). Functional fine-mapping analysis of this signal through a series of transgenic mouse reporter assays and gene rescue experiments led to the discovery of a novel growth enhancer of *GDF5*, named *GROW1*. A search for candidate alleles driving the positive selection signal surrounding *GROW1* identified the regulatory rs4911178-A allele, which is associated with short height, observed at a high frequency in non-African populations, and is present in Neanderthal and Denisovan genomes (Capellini et al. 2017). These findings highlight the potential of selection mapping in unmasking adaptive variants in regulatory elements of the genome.

Climatic adaptations: ultraviolet radiation Changes in exposure to ultraviolet radiation (UVR) accompanied changes in temperature as humans migrated out of Africa. The continuous spectrum of human skin color across the globe strongly maps onto the latitudinal distribution of UVR intensity (Chaplin 2004; Elias and Williams 2013; Jablonski and Chaplin 2000, 2013). Darker pigmentation is caused by an increased production of eumelanin in the dermis, which acts to absorb UV light and scatter it. In this way, melanin protects against dermal damage as well as UV-A photolysis of folate—the breakdown of which can lead to neural tube defects (Brenner and Hearing 2008; Kaidbey et al. 1979; Schmitz et al. 1995). Conversely, in regions of low UVR, lighter pigmentation is hypothesized to have been selectively favored for sufficient endogenous synthesis of Vitamin D, minimizing risk of Vitamin D deficiency (Loomis 1967).

Selection scans for candidate alleles underlying skin pigmentation adaptation have uncovered a handful of candidate genes that differ in signals between human populations, supporting evolutionary convergence (Alonso et al. 2008; de Gruijter et al. 2011; Izagirre et al. 2006; Lamason et al. 2005; Lao et al. 2007; Myles et al. 2007; Norton et al. 2007; Voight et al. 2006). For instance, evidence for positive selection in African populations has been found at pigmentation-associated genes *ASIP, MFSD12, DDB1, HERC2, TMEM138, OCA2, TP53BP1*, *DCT*, *TYRP1*, *EGFR,* and *DRD2* (Alonso et al. 2008; Crawford et al. 2017b; Izagirre et al. 2006; Lao et al. 2007), whereas *MC1R* is under strong functional constraint by purifying selection (John et al. 2003; Harding et al. 2000; Rana et al. 1999). In European populations, the strongest signals of positive selection have been found in *SLC24A5*, *TYRP1*, *KITLG*, *MATP,* and *TYR* (Lamason et al. 2005; Lao et al. 2007; Norton et al. 2007). *OCA2*, *DCT*, *KITLG*, *TYRP1*, *EGFR,* and *DRD2* have been suggested as candidates for positive selection in Asian populations (Alonso et al. 2008; Lao et al. 2007;

Myles et al. 2007). In indigenous American populations, 14 candidate genes under positive selection have been identified, including *SLC24A5, MATP, OPRM1,* and *EGFR* (Quillen et al. 2012). Within these candidate genes, several candidate pigmentation alleles have been fine-mapped and functionally annotated. Identified dark pigmentation alleles include: rs6058017-G (*ASIP*); rs6510760-A, rs10424065-T, rs112332856-C, and rs56203814-T (*MFSD12*); rs11230664-C (*DDB1*); rs7948623-T (*TMEM138*); rs1800404-C (*OCA2*); rs4932620-T and rs6497271-A (*HERC2*) (Crawford et al. 2017b). Associated light pigmentation alleles include: rs1426654-A (*SLC24A5*); rs16891982-G (*MATP*), rs6058017-A (*ASIP*); rs1800414-G and rs1448484-T (*OCA2*); rs1042602-A and rs1126809-A (*TYR*) (Lamason et al. 2005; Lao et al. 2007; Norton et al. 2007).

Recently, it has been suggested that light pigmentation alleles in modern human populations were present in the hominin lineage before modern humans evolved (Crawford et al. 2017b), supporting a model of selection on standing variation instead of convergent de novo variants in northern latitude groups. Significant associations with skin pigmentation variants in *SLC24A5, MFSD12, DDB1, TMEM138, OCA2,* and *HERC2* were observed among diverse African populations (Crawford et al. 2017b). The well-known light-pigmentation allele in *SLC24A5,* rs1426654-A, was introduced via gene flow from non-African groups into East Africa (Crawford et al. 2017b). The dark pigmentation alleles rs6510760-A, rs10424065-T, rs112332856-C, and rs56203814-T (*MFSD12*), rs11230664-C (*DDB1*), rs7948623-T (*TMEM138*), rs1800404-C (*OCA2*), rs4932620-T and rs6497271-A (*HERC2*) are identical by descent between African populations and populations from South Asia and Australo-Melanesia (Crawford et al. 2017b). These fascinating results demonstrate the evolutionary complexity of human pigmentation as the result of both shared ancestry and derived adaptive variation. Together, the extensive body of research on the genetics of human pigmentation variation represents one of the greatest selection mapping efforts across diverse human populations.

Dietary adaptations The advent of agriculture brought about new selective pressures on human populations in response to changes in diet. A prominent example of dietary adaptation is lactase persistence in diverse pastoralist populations. A 3.2 Mb region defining a strong signature of selection in these groups contains the lactase gene, *LCT* (Bersaglieri et al. 2004; Peter et al. 2012; Tishkoff et al. 2007; Voight et al. 2006). The protein encoded by this gene is responsible for breaking down lactose found in dairy products into glucose and galactose, thus enabling digestion in the intestines (Arribas et al. 2000). In northern European populations, two intronic SNPs upstream in *MCM6* (within the regulatory region of *LCT*) are under strong positive selection (rs4988235-T and rs182549-A) and regulates *LCT* expression (Enattah et al. 2002). Evidence supports recent de novo selection for these regulatory alleles (Bersaglieri et al. 2004; Peter et al. 2012; Tishkoff et al. 2007; Voight et al. 2006). Among African pastoralist populations from Tanzania, Kenya, and Sudan, three derived alleles (rs145946881-C, rs41525747-G, and rs41380347-G) in intron 13 of *MCM6* enhance *LCT* gene expression (Tishkoff et al. 2007). Lactase persistence thus represents another example of convergent evolution in humans from a shared cultural adaptation of adult milk consumption in northern Europe and East Africa.

Additional examples of dietary pressures driving human evolution include increased starch and fat consumption and exposure to arsenic in drinking water. Copy number variation of the salivary α-amylase gene (*AMY1*) is under positive selection among populations with dietary histories of high-starch consumption (Perry et al. 2007). Greater *AMY1* copy number corresponds to increased salivary amylase protein expression, which has led to the hypothesis that increased AMY1 protein levels enhances starch digestion and/or plays a role in reducing risk of intestinal disease (Perry et al. 2007). However, conflicting evidence exists for the biochemical role of AMY1 in starch digestion, and some scholars argue AMY1 expression profiles across tissues need to be considered when hypothesizing an evolutionary role (Fernandez and Wiley 2017). Selection mapping in other populations points to additional parts of the genome involved in starch digestion. Signals of positive selection in the maltase-glucoamylase gene, *MGAM,* involved in the final stages of starch digestion, have been identified from comparative analysis of ancient Andean genomes (~7000 years ago) and contemporary lowland and highland Peruvians (Lindo et al. 2018). The top candidate adaptive allele in *MGAM* is the C allele of intronic SNP rs77768615—a locus annotated with promoter-associated histone modifications in intestinal cells and enhancer-associated histone marks in cells of the duodenum, stomach, and intestines (Lindo et al. 2018). The findings suggest an adaptive shift to a starch-rich diet in these Peruvian populations consistent with archaeological evidence of intensive potato processing and consumption in the region (Haas et al. 2017; Rumold and Aldenderfer 2016).

Aside from food pressures, exceptionally high levels of the carcinogen arsenic have permeated the water source of the Colla population in the Puna region of Northwest Argentina for thousands of years (Eichstaedt et al. 2015). Candidate adaptive alleles rs1046778-C and rs7085104-G, located within the arsenic methyltransferase gene, *AS3MT,* under positive selection in the Colla population, are associated with decreased expression of *AS3MT* and lower excreted monomethylarsonic acid levels in urine (Eichstaedt et al. 2015; Schlebusch et al. 2013). Given that high levels of monomethylarsonic acid are associated with arsenic-related illness, these data suggest adaptive

variants altering *AS3MT* expression underlie adaptation to high arsenic environments in the Colla (Eichstaedt et al. 2015). Altogether, natural selection in response to dietary changes further exemplifies the diversity of adaptive genetic variation that underlies human adaptation.

Cardiovascular health Selection mapping is useful for the study of complex diseases, such as cardiovascular disease. As discussed earlier, when a trait is subjected to positive directional selection, linked variation rises in frequency along with the causative adaptive allele(s). In some cases, linked deleterious variants may exist on the adaptive haplotype(s). If the fitness benefits of the adaptive allele(s) outweigh the costs of the accompanied deleterious variants, those variants may persist in a population through time. In addition, beneficial (or neutral linked) alleles in a past environment may become deleterious alleles in a novel or changing environment, resulting in a gene–environment mismatch or interaction (Corbett et al. 2018). A powered phenotype-genotype association analysis of the top candidate loci under positive selection in a Mexican–American cohort identified candidate risk alleles for dyslipidemia—a highly prevalent condition among Amerindian-origin populations that results in elevated lipid levels in the blood (Ko et al. 2014). Among the top associated hits were well-known obesity and cardiovascular disease genes, previously associated risk alleles rs3135506-A and rs662799-C of *APOA5,* and novel risk alleles rs139961185-A of *SIK3* and intergenic rs28680850-A (nearby *LPL*) (Ko et al. 2014), potentially pointing to a gene-environment mismatch underlying disease risk in these populations. Similarly, it is well-described that high-altitude Andean populations, unlike other groups adapted to high altitude, exhibit polycythemia—an extreme production of red blood cells that can hinder blood flow to tissues from increased viscosity (León-Velarde 2003; Monge 1942). Top candidate alleles within haplotypes showing strong signatures of positive selection in a high-altitude Andean cohort—*BRINP3* (rs11578671-A), *NOS2* (rs34913965-T), and *TBX5* (rs10744822-C)—were associated phenotypic variation in cardiovascular health (Crawford et al. 2017a). The authors propose that high-altitude adaptation in Andeans involved changes in cardiovascular tolerance to polycythemia, instead of its reduction (Crawford et al. 2017a). Together, this work emphasizes how the application of selection mapping can aid in the discovery of risk alleles for complex disease, particularly when disease expression depends upon gene-environment interactions. Subsequent functional investigation of identified human candidate alleles for both adaptive traits and disease risk is necessary to determine the molecular mechanism, build models of human evolution, and facilitate medical research.

Downstream approaches for functional validation of candidate adaptive alleles

Functional investigation of candidate alleles under natural selection predicted to mediate adaptive phenotypic change is necessary to link putative adaptive biology to underlying molecular mechanisms. With the exception of forward genetics approaches noted above (e.g. malaria resistance alleles), most candidate adaptive alleles identified by evolutionary geneticists and molecular anthropologists have not been investigated functionally. The following discussion offers suggestions on how to pursue functional validation in cases where candidate adaptive alleles are identified, yet the molecular mechanism is unknown and thus the link to physiological adaptation is unclear. We include examples of functional investigations of human candidate adaptive alleles that provide supporting evidence for evolutionary hypotheses.

Myriad functional assays are currently available to investigate the contribution of high-priority candidate adaptive genetic variants to adaptive phenotypes. In order to devise an effective research strategy to test for putative adaptive roles, it is imperative that researchers confirm a genetic mechanism of action for the candidate adaptive allele(s), identify a biological readout of the adaptive allele(s), and develop model systems (Fig. 2). We divide this section into (1) functional assays to investigate genetic mechanism of candidate alleles, (2) systems to model molecular, cellular, and/or anatomical phenotypes, and (3) omics methods, which can collectively help link candidate adaptive variation to putative adaptive biology.

Functional assays for determining genetic mechanism

A major first step in detecting molecular phenotypes and ultimately linking this to adaptive biology is determining the genetic mechanism of candidate adaptive alleles. Genetic mechanism refers to the allele's effect on functional sequence, RNA, and protein products (Marchetti et al. 2012). As mentioned earlier, mining functional databases and published research can help guide functional assays, e.g., by identifying the cell type(s) in which the variant is highly expressed. There are numerous functional analyses that allow researchers to assess the genetic mechanism of candidate allele(s) on molecular biology, such as the effect on spatial and temporal expression profiles of RNA and protein across cell types and development, subcellular localization, splicing patterns, 3D structure, molecular interactions, and chromatin organization (Musunuru et al. 2018). Genetic variants that alter the protein sequence can have consequences for the protein's function, including loss-of-function alleles that reduce (hypomorph) or completely abrogate (null) protein activity, or gain-of-function alleles that confer novel or augmented (hypermorph) protein activity (Marchetti et al.

2012). For example, functional studies suggest both a gain-of-function and loss-of-function for alleles in *EGLN1* alleles among Tibetan highlanders, acting to lower hemoglobin concentration (Tashi et al. 2017; Song et al. 2020). Quantitative PCR (qPCR) is a commonly used technique to assess RNA expression of target RNAs in a sample (Maddocks and Jenkins 2017), which may offer preliminary evidence of alterations to RNA expression from the adaptive allele compared to the ancestral allele. This method uses reverse transcription of RNA to complementary DNA (cDNA) and absolute quantification of target molecules by calibration with DNA standard controls (Peirson and Butler 2007). If the candidate adaptive variant lies within a non-coding regulatory element (i.e., promoter, enhancer, silencer, insulator), gene reporter assays can be used to assess the impact of regulatory changes on gene expression (Davidson et al. 2011; Fu and Xiao 2006; Smemo et al. 2012). In this assay, the regulatory element is inserted within an expression plasmid upstream of a minimal promoter that drives expression of a detectable reporter gene, such as *lacZ* or luciferase (Davidson et al. 2011; Fu and Xiao 2006; Smemo et al. 2012). The ability and strength of the regulatory element to drive expression is then measured in cells transfected with the reporter construct via enzymatic activity of the reporter proteins (Davidson et al. 2011; Fu and Xiao 2006; Smemo et al. 2012). For example, fine-mapped, dark pigmentation-associated variants rs6510760-A and rs112332856-C, under selection in an upstream regulatory element of *MFSD12* in African populations, showed ~ 4.9 × decreased luciferase expression than light pigmentation-associated variants rs6510760-G and rs112332856-T (Crawford et al. 2017b). This reporter method can be modified for testing sequences predicted to affect splicing patterns by inserting the sequence between two exons and measuring expression of spliced transcripts in transfected cells using qPCR (Cooper 2005). Comparison of assays with wildtype and variant sequence can help assess allele-specific effects on splicing.

A generic approach for evaluating a genetic variant's effect on protein expression in a given biological sample is western blotting (Mahmood and Yang 2012). This technique first separates proteins by molecular weight through gel electrophoresis then transfers or blots proteins to a membrane where they are incubated with labeled protein-specific antibodies (Mahmood and Yang 2012). To determine if there is an impact on protein localization in the presence of the candidate adaptive variant, within tissues or specific cell-types, several techniques using protein- or RNA-specific labels have been developed, such as immunohistochemistry/immunocytochemistry and fluorescent in situ hybridization (FISH) (Coons et al. 1942; Langer-Safer et al. 1982). Cellular and subcellular localization also can be carried out using artificial peptide protein tags, such as a FLAG-tag or green fluorescent protein (GFP), genetically joined to the protein product via recombinant DNA techniques that require specific recognition antibodies for analysis (Einhauer and Jungbauer 2001; Kimple et al. 2013). The tag can then be visualized with immunofluorescence microscopy to reveal where the protein localizes (Einhauer and Jungbauer 2001; Kimple et al. 2013). Certain tags can be used in western blotting and immunoprecipitation, or to pull down the protein for 3D protein structure and X-ray crystallography analysis. Co-immunoprecipitation is a useful method for identifying altered or novel protein–protein interactions due to the variant using a protein-specific antibody to capture the target protein and any binding partners (Phizicky and Fields 1995). For every functional experiment, proper positive and negative controls as well as biological and technical replicates are crucial for validation. Steps to reduce experimental noise or off-target effects need to be taken to ensure the data reflect true molecular changes. After gaining preliminary insight into the basic molecular mechanism of the candidate adaptive allele, it is necessary to perform further experiments using appropriate model systems to delineate the contribution of the candidate adaptive alleles to putative adaptive biology (cellular, anatomical, physiological), and to inform evolutionary models of adaptation.

Modeling adaptive phenotypes

It is well known from work in molecular biology that the expression of specific RNA transcripts and protein isoforms, or use of certain regulatory elements, often exhibit developmental and cell-type specificity across organisms (Cardoso-Moreira et al. 2019; Mayer et al. 2019; Pollen et al. 2019; Sapkota et al. 2019). In vitro and in vivo model systems therefore offer a great opportunity to capture or enrich for the relevant, affected biology of the candidate variant predicted to underlie physiological adaptations. These model systems include bacteria, yeast, animal models, and cell culture. Model organisms such as the mouse, zebrafish, ferret, macaque, fruit fly, and *C. elegans* are commonly used (Leonelli and Ankeny 2013), depending on the ability of that organism to recapitulate the human phenotype(s) of interest, feasibility of genetic manipulation, and generational time course. In recent years, some studies have modified the genomes of animal models for functional characterization of candidate adaptive alleles and define their putative role in human evolution. Transgenic mouse models carrying the candidate derived regulatory allele (rs4911178-A) in the *GROW1* enhancer of *GDF5* for human height adaptation showed significantly reduced long bone length compared to the ancestral allele (Capellini et al. 2017). Additionally, a human-specific deletion of a regulatory enhancer for penile spine development, likely from selection relaxation in humans, results in failure of penile spine formation when knocked out in mouse models (McLean et al. 2011).

In conjunction with model organisms, researchers often use in vitro cell lines (Carter and Shieh 2015). Cell culturing is favorable as it offers a platform to model human cellular biology, and can offer a feasible, malleable, and quick way to obtain informative functional data (Carter and Shieh 2015). For example, cultured human keratinocytes were used to investigate the impact of a candidate adaptive SNP in *KITLG* under positive selection in Europeans and associated with blond hair color phenotypes, revealing an alteration in the LEF1 transcription factor binding site coupled with a reduction in enhancer activity (Guenther et al. 2014).

Cell lines for culturing can be obtained directly from primary tissue, or derived from an established cell line or strain. For instance, stem cell lines can be reprogrammed and differentiated into different cell types, or tissues, in the body (Park et al. 2008). Stem cells can either be embryonically-obtained or generated from chemically-inducing reprogrammable adult cells, such as fibroblasts, to a pluripotent stem cell-like state (Chin et al. 2009; Zhou and Ding 2010). The latter are referred to as induced pluripotent stem cells (iPSCs). Recent advancements in stem cell technology allow for the differentiation of iPSCs into 3D structures called 'organoids' that recapitulate in vivo organs, dependent on a standardized combination of nutrients, growth factors, inhibitors, and other small molecules to specify organogenesis (Yin et al. 2016). This model system may be especially informative if the associated adaptive biology of the candidate allele is developmental (Lancaster et al. 2013). Perhaps the greatest insights from organoid models have come from modeling species-level adaptive variation in brain development (Kanton et al. 2019; Li et al. 2017; Pollen et al. 2019). For instance, recent work using cerebral organoids has demonstrated that human-specific *NOTCH2NL* gene duplicates—candidates for human brain evolution—are highly expressed in neural progenitor cells in the developing brain (Fiddes et al. 2018; Suzuki et al. 2018). Additionally, their expression is involved in delaying the differentiation of these cells, which corresponds to an increase in neuronal output underlying a larger brain (Fiddes et al. 2018; Suzuki et al. 2018). While this work shows the utility of organoid modeling of macroevolutionary differences, one can imagine the use of organoids for modeling putative adaptive alleles within human populations (e.g., spleen organoids to model the candidate *PDE10A* SNP for deep-sea diving adaptation (Ilardo et al. 2018)). The appropriate model system is imperative for the applicability of functional findings to the understanding of human evolution. For example, using a mouse model to explore the role of a candidate genomic region during human brain development may be less useful than using human fetal brain tissue and/or human cerebral organoids as a result of unique features of cellular proliferation, differentiation, and organization of the human brain that are not conserved in the mouse (Huch and Koo 2015).

Upon selecting an appropriate model system, knockdown, knockout, and overexpression assays can be used to test sufficiency and/or necessity of the gene or regulatory region containing the variant for a particular biological feature, or to test the downstream targets of the variant, if not previously conducted. These data can help guide additional phenotyping. Such methods include CRISPR-Cas9 knockout (Ran et al. 2013), morpholino knockdown (Timme-Laragy et al. 2012), RNA interference (Agrawal et al. 2003), and CRISPR gene knock-in into a safe harbor locus (Verma et al. 2017). For example, morpholino knockdown of *SLC24A5* in zebrafish (Lamason et al. 2005) and CRISPR-Cas9 knockout of *MFSD12* in mice (Crawford et al. 2017b)—two genes under positive selection in humans and harboring candidate pigmentation alleles—confirmed a functional role in melanogenesis.

After selection of a suitable model system, one must determine a method for introducing the candidate allele. Given that the genetic variants may be as small as a SNV or as large as a chromosomal aberration, efficacy and feasibility must be considered. Two sources for which the desired allele can be acquired include: (1) original genomic sequence from the individual(s) with the allele, or (2) genetic engineering (Fig. 2) (White 1997). Aside from primary tissue or iPSC-based analyses, cloning of the human genetic variants into a vector system for transfection may be desired for functional assays in bacteria, yeast, or in vivo transgenic models (Reed et al. 2017). For molecular cloning, the target region of interest is amplified and ligated into a selected cloning vector, generating recombinant DNA that can be introduced into the host organism for replication (Ashwini et al. 2016). Standard vectors contain a DNA replication origin, unique restriction endonuclease recognition sites for introduction of selectable marker gene(s) that aid in the selection of cells with successful integration of the vector sequence, and a tag gene for screening (Ashwini et al. 2016). In addition, if it is desired to have control over when the variant is expressed during development, or in which cell types, one might use a cell-type specific promoter or chemically inducible promoter systems, such as the piggy-bac transposon inducible expression system (Glover et al. 2013).

Artificially introducing genetic variation depends on the type of genetic variant. For instance, genetically engineering a bi-allelic substitution into the model system requires a different methodology than introducing a heterozygous deletion (Supharattanasitthi et al. 2019). The most recent and effective technique in gene editing is the CRISPR/Cas9 system in which guide RNAs direct the Cas9 nuclease to a specific genomic site and induce a double-stranded break (Cong et al. 2013). CRISPR/Cas9 is most often used to knockout a gene by introducing a deletion at the target site. However, current technological advancements allow for single base substitution changes by use of single-stranded

oligodeoxynucleotides that are tethered to the Cas9-guide RNA ribonucleoprotein complex to promote homology-directed repair (Aird et al. 2018; Zhang et al. 2019).

Omics technologies

Downstream 'omics' approaches can offer valuable information on the global impact of the candidate adaptive variant(s) on expression dynamics and/or interaction networks involving several biological pathways across cells and tissues that underlie modeled phenotypes (Manzoni et al. 2018). These approaches include transcriptomics (RNA), proteomics (protein), and interactomics (molecular interactions) (Manzoni et al. 2018). Research studying human variation and disease is growing towards the integration of different omics techniques to concurrently evaluate different aspects of molecular biology (Karczewski and Snyder 2018). This exciting era of integrative omics analyses offers novel opportunities to investigate complex biological processes underlying adaptive evolution.

The complete set of transcribed or expressed RNA within a cell, tissue, or whole organism is referred to as the transcriptome (Berg 2001). Distinct features of expressed RNA can be exploited experimentally to capture the transcriptome of a sample using a method known as transcriptomics (Lowe et al. 2017). A popular method is bulk RNA-seq, a high-throughput sequencing approach that extracts and reverse transcribes RNA molecules from a population of cells in a given sample together (i.e., 'bulk), generating a cDNA library for sequencing, processing, and alignment to a reference for averaged quantification of transcript expression within the sample (Wang et al. 2009). Library preparation for bulk RNA-seq can be performed to include the complex suite of functional RNA molecules present in the sample (including ribosomal RNA, transfer RNA, non-coding RNAs, and messenger RNA (mRNA)), or enriched for protein-coding mRNAs through polyadenylation capture techniques and/or depletion of non-mRNAs (Lowe et al. 2017; Wang et al. 2009; Zhao et al. 2014). Upregulation and downregulation of genes and gene sets in a sample positive for the candidate adaptive allele compared to wildtype can be assessed using bioinformatics analyses. For example, RNA-seq of cultured primary melanocytes obtained from African individuals carrying derived upstream regulatory alleles of *MFSD12* (rs6510760-A and rs112332856-C) confirmed a significantly decreased expression of *MFSD12* compared to cultured melanocytes derived from individuals carrying the ancestral alleles (Crawford et al. 2017b).

More recently, methods have been developed to obtain sequencing data at single-cell resolution, allowing for the potential discovery of novel and/or unexpected biology not captured with traditional bulk sample approaches (Chen et al. 2019; Gierahn et al. 2017; Islam et al. 2014; Jaitin et al. 2014; Macosko et al. 2015). Single-cell RNA sequencing (scRNA-seq) methods resolve cellular heterogeneity in a sample, allow for cell lineage tracing during development, can uncover novel or rare cell populations, and often inform gene regulatory relationships (Hedlund and Deng 2018). The general analysis pipeline begins with sample dissociation, single-cell isolation, cell and transcript barcoding, library preparation, and high-throughput sequencing (Chen et al. 2019). Single-cell isolation can be carried out through the use of cell sorting techniques or through micro-well arrays, and transcripts are often captured via the 3′ end polyadenylation tail common to many protein-coding genes for reverse transcription (Gierahn et al. 2017; Islam et al. 2014; Jaitin et al. 2014; Macosko et al. 2015). Following sequencing, quality control and alignment to a reference genome are carried out, and data is often analyzed using unsupervised graph-based clustering methods, such as the Seurat R toolkit pipeline (Butler et al. 2018). Given proper control of batch effects, scRNA-seq comparison across genotypes can expose differentially expressed genes within an affected cell population of a sample, changes in cellular composition, and/or variation in cellular differentiation rates, together providing insight into allele-specific functional effects (Haghverdi et al. 2018). Current efforts in the field are focused on improving the ability to comprehensively integrate single-cell sequencing data across experiments and modalities (Stuart et al. 2019).

Other approaches seek to evaluate the entire set of translated proteins within a cell, tissue, or sample (the proteome) called proteomics. Two-dimensional gel electrophoresis combined with mass spectrometry, or more recently shotgun mass spectrometry, is often used for the identification and quantification of the sample proteome (Martins-de-Souza 2014). Proteome studies tend to combine approaches that characterize other features of proteins, such as post-translational modifications or interacting partners (Ke et al. 2016). Other research methods seek an ample readout of the molecular interaction dynamics in a sample, such as DNA-DNA, DNA-RNA, DNA-protein, and/or protein-protein—collectively called interactomics (Salas et al. 2020). A recent protein-protein interaction assay, Affinity Capture of Polyribosomes followed by RNA sequencing (ACAPseq) captures and sequences the polyribosomes of nascent proteins that bind to a bait protein to detect protein interactions (Peng et al. 2018). Thus, this approach may be very useful to assess changes in protein interactions for a candidate adaptive variant that alters a specific protein domain. Changes on genome-wide protein interactions with DNA, as well as epigenetic histone and DNA modifications, can be assessed with chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Raha et al. 2010). One may wish to assess genome-wide changes in such interactions if the candidate adaptive variant affects the activity of a known chromatin

modifier or a transcription factor, for example. There are also recently optimized single-cell ChIP-seq pipelines available that offer better resolution of the heterogeneity of chromatin states across cell-types (Grosselin et al. 2019).

Candidate adaptive variants (especially deletions, duplications, inversions, and translocations) may disrupt or re-organize topologically-associated domains (TADs)—domains that are proposed to partition the genome into regulatory units wherein several enhancers and promoters interact (Szabo et al. 2019). Chromatin conformation capture methods, such as Hi-C, evaluate spatial organization of chromatin in the cell and can be useful for understanding genomic changes that alter TAD structures (Lieberman-Aiden et al. 2009). Pairing findings of TAD disruption or re-organization with changes in expression can thus offer substantial insight into the mechanism of action of adaptive alleles. In general, there are limitations to be considered for each omics approach, in terms of method efficacy, batch effects, bioinformatics, and interpretation of results. Often, these results are hypothesis-generating and require additional follow-up experiments using specific functional assays. The most robust arguments for putative functional roles of candidate adaptive alleles will build on data from several model systems and experiments to generate sufficient biological readouts for confirmation and validation. Thorough and systematic functional validation work is crucial for informing adaptive hypotheses and models of human evolution.

Conclusion

Selection mapping can be informative for the broader scope of human biology, health, and disease when combined with association, fine-mapping, and functional analyses. Given the complexity and diversity of human genetic variation, environmental and cultural conditions, and gene-culture-environment interactions across geographicallydefined human populations through time, it is crucial for scientists to increase the number of selection mapping studies in non-European and admixed populations in order to better understand how natural selection shapes human biology across varying ancestral genomic backgrounds (Bentley et al. 2017, 2020; Popejoy and Fullerton 2016; Sirugo et al. 2019). It is imperative to note that scientific bias in human genomic sciences is not limited to which populations are being studied. Greater efforts must also be made to increase representation within the community of scientific leaders who study human genetic variation beyond westernized scientists to include individuals from under-represented, vulnerable, and indigenous populations (Claw et al. 2018; Jackson et al. 2019). A more representative, global map of functional variants under natural selection in humans will have translational value for variant annotation in clinical disease cohorts, which are currently barraged with hundreds of variants of unknown significance from whole genome and whole exome sequencing data. In addition, given the limited sensitivity of current genotype–phenotype methods to polygenic traits or rare variants, selection mapping can enhance the power of these analyses to detect causal variants. Nonetheless, limitations of current selection detection methods need to be overcome to inform the genetics of complex traits, and include improving both statistical inference methods and theory to better model complex soft selective sweeps. Moving forward, inclusion of additional global populations, particularly those with unique adaptive histories, to the growing database of selection maps will provide a fruitful source for phenotypic and disease mapping. Pairing selection mapping with association analysis has been useful for fine mapping causal variants of several traits in humans and other species. Beyond genotype–phenotype mapping, genome-wide selection mapping has the potential to enrich epigenetic, transcriptomic, proteomic, and interactomic research by pointing to functional factors that are evolutionarily conserved or adaptable.

Acknowledgements We want to first and foremost thank all the study participants from diverse populations across the globe—and through time—who were included in the research papers reviewed here. We would like to thank Stephanie Bielas for providing helpful comments on the manuscript and Michael Bamshad for early ideas related to selection mapping.

Funding This research has no funding source.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA, Consortium GP (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073. https://doi.org/10.1038/nature09534

Agrawal N, Dasaradhi PV, Mohmmed A, Malhotra P, Bhatnagar RK, Mukherjee SK (2003) RNA interference: biology, mechanism, and applications. Microbiol Mol Biol Rev 67:657–685. https://doi.org/10.1128/mmbr.67.4.657-685.2003

Aird EJKNL, Martin AS, Harris RS, Gordon WR (2018) Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. Commun Biol 1:54

Akbari A, Vitti JJ, Iranmehr A, Bakhtiari M, Sabeti PC, Mirarab S, Bafna V (2018) Identifying the favored mutation in a positive selective sweep. Nat Methods 15:279–282. https://doi.org/10.1038/nmeth.4606

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12:1805–1814. https://doi.org/10.1101/gr.631202

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol 2:e286. https://doi.org/10.1371/journal.pbio.0020286

Algady W, Louzada S, Carpenter D, Brajer P, Farnert A, Rooth I, Ngasala B, Yang F, Shaw MA, Hollox EJ (2018) The Malaria-protective human glycophorin structural variant DUP4 shows somatic mosaicism and association with hemoglobin levels. Am J Hum Genet 103:769–776. https://doi.org/10.1016/j.ajhg.2018.10.008

Alkorta-Aranburu G, Beall CM, Witonsky DB, Gebremedhin A, Pritchard JK, Di Rienzo A (2012) The genetic architecture of adaptations to high altitude in Ethiopia. PLoS Genet 8:e1003110. https://doi.org/10.1371/journal.pgen.1003110

Allison AC (1954) Notes on sickle-cell polymorphism. Ann Hum Genet 19:39–51. https://doi.org/10.1111/j.1469-1809.1954.tb01261.x

Alonso S, Izagirre N, Smith-Zubiaga I, Gardeazabal J, Diaz-Ramon JL, Diaz-Perez JL, Zelenika D, Boyano MD, Smit N, de la Rua C (2008) Complex signatures of selection for the melanogenic loci TYR, TYRP1 and DCT in humans. BMC Evol Biol 8:74. https://doi.org/10.1186/1471-2148-8-74

Amato R, Miele G, Monticelli A, Cocozza S (2011) Signs of selective pressure on genetic variants affecting human height. PLoS ONE 6:e27588. https://doi.org/10.1371/journal.pone.0027588

Andolfatto P (2001) Adaptive hitchhiking effects on genome variability. Curr Opin Genet Dev 11:635–641

Arribas JCD, Herrero AG, Martín-Lomas M, Cañada FJ, He S, Withers SG (2000) Differential mechanism-based labeling and unequivocal activityassignment of the two active sites of intestinal lactase/phlorizinhydrolase. Eur J Biochem 267:6996–7005

Ashwini M, Murugan SB, Balamurugan S, Sathishkumar R (2016) Advances in Molecular Cloning. Mol Biol (Mosk) 50:3–9. https://doi.org/10.7868/S0026898416010134

Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111. https://doi.org/10.1038/nrg999

Bamshad MJ, Mummidi S, Gonzalez E, Ahuja SS, Dunn DM, Watkins WS, Wooding S, Stone AC, Jorde LB, Weiss RB, Ahuja SK (2002) A strong signature of balancing selection in the 5' cis-regulatory region of CCR5. Proc Natl Acad Sci USA 99:10539–10544. https://doi.org/10.1073/pnas.162046399

Barrett RD, Schluter D (2008) Adaptation from standing genetic variation. Trends Ecol Evol 23:38–44. https://doi.org/10.1016/j.tree.2007.09.008

Basu Mallick C, Iliescu FM, Möls M, Hill S, Tamang R, Chaubey G, Goto R, Ho SY, Gallego Romero I, Crivellaro F, Hudjashov G, Rai N, Metspalu M, Mascie-Taylor CG, Pitchappan R, Singh L, Mirazon-Lahr M, Thangaraj K, Villems R, Kivisild T (2013) The light skin allele of SLC24A5 in South Asians and Europeans shares identity by descent. PLoS Genet 9:e1003912. https://doi.org/10.1371/journal.pgen.1003912

Beall CM (2013) Human adaptability studies at high altitude: research designs and major concepts during fifty years of discovery. Am J Hum Biol 25:141–147. https://doi.org/10.1002/ajhb.22355

Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, Li C, Li JC, Liang Y, McCormack M, Montgomery HE, Pan H, Robbins PA, Shianna KV, Tam SC, Tsering N, Veeramah KR, Wang W, Wangdui P, Weale ME, Xu Y, Xu Z, Yang L, Zaman MJ, Zeng C, Zhang L, Zhang X, Zhaxi P, Zheng YT (2010) Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. Proc Natl Acad Sci USA 107:11459–11464. https://doi.org/10.1073/pnas.1002443107

Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. Nature 356:519–520. https://doi.org/10.1038/356519a0

Bentley AR, Callier S, Rotimi CN (2017) Diversity and inclusion in genomic research: why the uneven progress? J Community Genet. https://doi.org/10.1007/s12687-017-0316-6

Bentley AR, Callier SL, Rotimi CN (2020) Evaluating the promise of inclusion of African ancestry populations in genomics. NPJ Genom Med 5:5. https://doi.org/10.1038/s41525-019-0111-x

Berg JM (2001) Biochemistry. Chem Eng News 79:130–130

Berg JJ, Coop G (2014) A population genetic signal of polygenic adaptation. PLoS Genet 10:e1004412. https://doi.org/10.1371/journal.pgen.1004412

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN (2004) Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111–1120. https://doi.org/10.1086/421051

Bigham AW, Lee FS (2014) Human high-altitude adaptation: forward genetics meets the HIF pathway. Genes Dev 28:2189–2204. https://doi.org/10.1101/gad.250167.114

Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R, Scherer SW, Julian CG, Wilson MJ, Lopez Herraez D, Brutsaert T, Parra EJ, Moore LG, Shriver MD (2010) Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genet 6:e1001116. https://doi.org/10.1371/journal.pgen.1001116

Bigham AW, Buckingham KJ, Husain S, Emond MJ, Bofferding KM, Gildersleeve H, Rutherford A, Astakhova NM, Perelygin AA, Busch MP, Murray KO, Sejvar JJ, Green S, Kriesel J, Brinton MA, Bamshad M (2011) Host genetic risk factors for West Nile virus infection and disease progression. PLoS ONE 6:e24745. https://doi.org/10.1371/journal.pone.0024745

Bigham AW, Julian CG, Wilson MJ, Vargas E, Browne VA, Shriver MD, Moore LG (2014) Maternal PRKAA1 and EDNRA genotypes are associated with birth weight, and PRKAA1 with uterine artery diameter and metabolic homeostasis at high altitude. Physiol Genomics 46:687–697

Bigham AW, Magnaye K, Dunn DM, Weiss RB, Bamshad M (2018) Complex signatures of natural selection at GYPA. Hum Genet 137:151–160. https://doi.org/10.1007/s00439-018-1866-3

Biswas S, Akey JM (2006) Genomic insights into positive selection. Trends Genet 22:437–446

Brenner M, Hearing VJ (2008) The protective role of melanin against UV damage in human skin. Photochem Photobiol 84:539–549. https://doi.org/10.1111/j.1751-1097.2007.00226.x

Brutsaert TD, Kiyamu M, Revollendo GE, Isherwood JL, Lee FS, Rivera-Ch M, Leon-Velarde F, Ghosh S, Bigham AW (2019) Association of EGLN1 gene with high aerobic capacity of Peruvian Quechua at high altitude. Proc Natl Acad Sci 116:24006–24011

Butler A, Hoffman P, Smibert P, Papalexi E, Satija R (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36:411–420. https://doi.org/10.1038/nbt.4096

Buxton P, Edwards C, Archer CW, Francis-West P (2001) Growth/differentiation factor-5 (GDF-5) and skeletal development. J Bone Jt Surg Am 83(Suppl 1):S23–30

Camus D, Hadley TJ (1985) A Plasmodium falciparum antigen that binds to host erythrocytes and merozoites. Science 230:553–556. https://doi.org/10.1126/science.3901257

Capellini TD, Chen H, Cao J, Doxey AC, Kiapour AM, Schoor M, Kingsley DM (2017) Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. Nat Genet 49:1202–1210. https://doi.org/10.1038/ng.3911

Cardona A, Pagani L, Antao T, Lawson DJ, Eichstaedt CA, Yngvadottir B, Shwe MT, Wee J, Romero IG, Raj S, Metspalu M, Villems R, Willerslev E, Tyler-Smith C, Malyarchuk BA, Derenko MV, Kivisild T (2014) Genome-wide analysis of cold adaptation in

indigenous Siberian populations. PLoS ONE 9:e98076. https://doi.org/10.1371/journal.pone.0098076
Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, Liechti A, Ascenção K, Rummel C, Ovchinnikova S, Mazin PV, Xenarios I, Harshman K, Mort M, Cooper DN, Sandi C, Soares MJ, Ferreira PG, Afonso S, Carneiro M, Turner JMA, VandeBerg JL, Fallahshahroudi A, Jensen P, Behr R, Lisgo S, Lindsay S, Khaitovich P, Huber W, Baker J, Anders S, Zhang YE, Kaessmann H (2019) Gene expression across mammalian organ development. Nature 571:505–509. https://doi.org/10.1038/s41586-019-1338-5
Carter M, Shieh J (2015) Chapter 14-cell culture techniques. Guide to research techniques in neuroscience. Elsevier, Amsterdam, pp 295–310
Cavalli-Sforza LL (1966) Population structure and human evolution. Proc R Soc Lond B Biol Sci 164:362–379
Chaplin G (2004) Geographic distribution of environmental factors influencing human skin coloration. Am J Phys Anthropol 125:292–302. https://doi.org/10.1002/ajpa.10263
Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303
Chen H, Patterson N, Reich D (2010) Population differentiation as a test for selective sweeps. Genome Res 20:393–402. https://doi.org/10.1101/gr.100545.109
Chen G, Ning B, Shi T (2019) Single-cell RNA-Seq technologies and related computational data analysis. Front Genet 10:317. https://doi.org/10.3389/fgene.2019.00317
Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, Ambartsumyan G, Aimiuwu O, Richter L, Zhang J, Khvorostov I, Ott V, Grunstein M, Lavon N, Benvenisty N, Croce CM, Clark AT, Baxter T, Pyle AD, Teitell MA, Pelegrini M, Plath K, Lowry WE (2009) Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell 5:111–123. https://doi.org/10.1016/j.stem.2009.06.008
Cholera R, Brittain NJ, Gillrie MR, Lopera-Mesa TM, Diakite SA, Arie T, Krause MA, Guindo A, Tubman A, Fujioka H, Diallo DA, Doumbo OK, Ho M, Wellems TE, Fairhurst RM (2008) Impaired cytoadherence of *Plasmodium falciparum*-infected erythrocytes containing sickle hemoglobin. Proc Natl Acad Sci USA 105:991–996. https://doi.org/10.1073/pnas.0711401105
Chowdhury F, Rahman MA, Begum YA, Khan AI, Faruque AS, Saha NC, Baby NI, Malek MA, Kumar AR, Svennerholm AM, Pietroni M, Cravioto A, Qadri F (2011) Impact of rapid urbanization on the rates of infection by Vibrio cholerae O1 and enterotoxigenic Escherichia coli in Dhaka. Bangladesh PLoS Negl Trop Dis 5:e999. https://doi.org/10.1371/journal.pntd.0000999
Claw KG, Anderson MZ, Begay RL, Tsosie KS, Fox K, Garrison NA, Consortium SifIpiGS (2018) A framework for enhancing ethical genomic research with Indigenous communities. Nat Commun 9:2957. https://doi.org/10.1038/s41467-018-05188-3
Clemente FJ, Cardona A, Inchley CE, Peter BM, Jacobs G, Pagani L, Lawson DJ, Antao T, Vicente M, Mitt M, DeGiorgio M, Faltyskova Z, Xue Y, Ayub Q, Szpak M, Magi R, Eriksson A, Manica A, Raghavan M, Rasmussen M, Rasmussen S, Willerslev E, Vidal-Puig A, Tyler-Smith C, Villems R, Nielsen R, Metspalu M, Malyarchuk B, Derenko M, Kivisild T (2014) A Selective sweep on a deleterious mutation in CPT1A in Arctic populations. Am J Hum Genet 95:584–589. https://doi.org/10.1016/j.ajhg.2014.09.016
Collins SA, Sinclair G, McIntosh S, Bamforth F, Thompson R, Sobol I, Osborne G, Corriveau A, Santos M, Hanley B, Greenberg CR, Vallance H, Arbour L (2010) Carnitine palmitoyltransferase 1A (CPT1A) P479L prevalence in live newborns in Yukon, Northwest Territories, and Nunavut. Mol Genet Metab 101:200–204. https://doi.org/10.1016/j.ymgme.2010.07.013
Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823. https://doi.org/10.1126/science.1231143
Coons AHCHJ, Jones RN, Berliner E (1942) The demonstration of pneumococcal antigen in tissues by the use of fluorescent antibody. J Immunol 45:159–170
Cooper TA (2005) Use of minigene systems to dissect alternative splicing elements. Methods 37:331–340. https://doi.org/10.1016/j.ymeth.2005.07.015
Corbett S, Courtiol A, Lummaa V, Moorad J, Stearns S (2018) The transition to modernity and chronic disease: mismatch and natural selection. Nat Rev Genet 19:419–430. https://doi.org/10.1038/s41576-018-0012-3
Crawford JE, Amaru R, Song J, Julian CG, Racimo F, Cheng JY, Guo X, Yao J, Ambale-Venkatesh B, Lima JA, Rotter JI, Stehlik J, Moore LG, Prchal JT, Nielsen R (2017a) Natural Selection on Genes Related to Cardiovascular Health in High-Altitude Adapted Andeans. Am J Hum Genet 101:752–767. https://doi.org/10.1016/j.ajhg.2017.09.023
Crawford NG, Kelly DE, Hansen MEB, Beltrame MH, Fan S, Bowman SL, Jewett E, Ranciaro A, Thompson S, Lo Y, Pfeifer SP, Jensen JD, Campbell MC, Beggs W, Hormozdiari F, Mpoloka SW, Mokone GG, Nyambo T, Meskel DW, Belay G, Haut J, Program NCS, Rothschild H, Zon L, Zhou Y, Kovacs MA, Xu M, Zhang T, Bishop K, Sinclair J, Rivas C, Elliot E, Choi J, Li SA, Hicks B, Burgess S, Abnet C, Watkins-Chow DE, Oceana E, Song YS, Eskin E, Brown KM, Marks MS, Loftus SK, Pavan WJ, Yeager M, Chanock S, Tishkoff SA (2017b) Loci associated with skin pigmentation identified in African populations. Science. https://doi.org/10.1126/science.aan8433
Cvijovic I, Good BH, Desai MM (2018) The effect of strong purifying selection on genetic diversity. Genetics 209:1235–1278. https://doi.org/10.1534/genetics.118.301058
Daanen HA, Van Marken Lichtenbelt WD (2016) Human whole body cold adaptation. Temperature (Austin) 3:104–118. https://doi.org/10.1080/23328940.2015.1135688
Davidson S, Lear M, Shanley L, Hing B, Baizan-Edge A, Herwig A, Quinn JP, Breen G, McGuffin P, Starkey A, Barrett P, MacKenzie A (2011) Differential activity by polymorphic variants of a remote enhancer that supports galanin expression in the hypothalamus and amygdala: implications for obesity, depression and alcoholism. Neuropsychopharmacology 36:2211–2221. https://doi.org/10.1038/npp.2011.93
de Gruijter JM, Lao O, Vermeulen M, Xue Y, Woodwark C, Gillson CJ, Coffey AJ, Ayub Q, Mehdi SQ, Kayser M, Tyler-Smith C (2011) Contrasting signals of positive selection in genes involved in human skin-color variation from tests based on SNP scans and resequencing. Investig Genet 2:24. https://doi.org/10.1186/2041-2223-2-24
Eaaswarkhanth M, Dos Santos ALC, Gokcumen O, Al-Mulla F, Thanaraj TA (2020) Genome-wide selection scan in an Arabian Peninsula population identifies a TNKS haplotype linked to metabolic traits and hypertension. Genome Biol Evol 12:77–87. https://doi.org/10.1093/gbe/evaa033
Eichstaedt CA, Antao T, Cardona A, Pagani L, Kivisild T, Mormina M (2015) Positive selection of AS3MT to arsenic water in Andean populations. Mutat Res 780:97–102. https://doi.org/10.1016/j.mrfmmm.2015.07.007
Eichstaedt CA, Pagani L, Antao T, Inchley CE, Cardona A, Morseburg A, Clemente FJ, Sluckin TJ, Metspalu E, Mitt M, Magi R, Hudjashov G, Metspalu M, Mormina M, Jacobs GS, Kivisild T (2017) Evidence of early-stage selection on EPAS1 and GPR126 genes in andean high altitude populations. Sci Rep 7:13042. https://doi.org/10.1038/s41598-017-13382-4

Einhauer A, Jungbauer A (2001) The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. J Biochem Biophys Methods 49:455–465. https://doi.org/10.1016/s0165-022x(01)00213-5

Elias PM, Williams ML (2013) Re-appraisal of current theories for the development and loss of epidermal pigmentation in hominins and modern humans. J Hum Evol 64:687–692. https://doi.org/10.1016/j.jhevol.2013.02.003

Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I (2002) Identification of a variant associated with adult-type hypolactasia. Nat Genet 30:233–237. https://doi.org/10.1038/ng826

Evans AG, Wellems TE (2002) Coevolutionary genetics of Plasmodium malaria parasites and their human hosts. Integr Comp Biol 42:401–407. https://doi.org/10.1093/icb/42.2.401

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Popul Biol 3:87–112. https://doi.org/10.1016/0040-5809(72)90035-4

Faruque SM, Sack DA, Sack RB, Colwell RR, Takeda Y, Nair GB (2003) Emergence and evolution of Vibrio cholerae O139. Proc Natl Acad Sci USA 100:1304–1309. https://doi.org/10.1073/pnas.0337468100

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Fehren-Schmitz L, Georges L (2016) Ancient DNA reveals selection acting on genes associated with hypoxia response in pre-Columbian Peruvian Highlanders in the last 8500 years. Sci Rep 6:23485. https://doi.org/10.1038/srep23485

Feldman MW, Cavalli-Sforza LL (1976) Cultural and biological evolutionary processes, selection for a trait under complex transmission. Theor Popul Biol 9:238–259. https://doi.org/10.1016/0040-5809(76)90047-2

Fernandez CI, Wiley AS (2017) Rethinking the starch digestion hypothesis for AMY1 copy number variation in humans. Am J Phys Anthropol 163:645–657. https://doi.org/10.1002/ajpa.23237

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R (2014) On detecting incomplete soft or hard selective sweeps using haplotype structure. Mol Biol Evol 31:1275–1291. https://doi.org/10.1093/molbev/msu077

Fiddes IT, Lodewijk GA, Mooring M, Bosworth CM, Ewing AD, Mantalas GL, Novak AM, van den Bout A, Bishara A, Rosenkrantz JL, Lorig-Roach R, Field AR, Haeussler M, Russo L, Bhaduri A, Nowakowski TJ, Pollen AA, Dougherty ML, Nuttle X, Addor MC, Zwolinski S, Katzman S, Kriegstein A, Eichler EE, Salama SR, Jacobs FMJ, Haussler D (2018) Human-specific NOTCH2NL genes affect notch signaling and cortical neurogenesis. Cell 173(1356–1369):e22. https://doi.org/10.1016/j.cell.2018.03.051

Fischer A, Rausell A (2016) Primary immunodeficiencies suggest redundancy within the human immune system. Sci Immunol 1:eaah5861. https://doi.org/10.1126/sciimmunol.aah5861

Francis-West PH, Parish J, Lee K, Archer CW (1999) BMP/GDF-signalling interactions during synovial joint development. Cell Tissue Res 296:111–119. https://doi.org/10.1007/s004410051272

Fu YX (1995) Statistical properties of segregating sites. Theor Popul Biol 48:172–197. https://doi.org/10.1006/tpbi.1995.1025

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Fu Y, Xiao W (2006) Study of transcriptional regulation using a reporter gene assay. Methods Mol Biol 313:257–264. https://doi.org/10.1385/1-59259-958-3:257

Fumagalli M, Pozzoli U, Cagliani R, Comi GP, Bresolin N, Clerici M, Sironi M (2010) Genome-wide identification of susceptibility alleles for viral infections through a population genetics approach. PLoS Genet 6:e1000849. https://doi.org/10.1371/journal.pgen.1000849

Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Ferrer-Admettla A, Pattini L, Nielsen R (2011) Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet 7:e1002355. https://doi.org/10.1371/journal.pgen.1002355

Galvani AP, Slatkin M (2003) Evaluating plague and smallpox as historical selective pressures for the CCR5-Delta 32 HIV-resistance allele. Proc Natl Acad Sci USA 100:15276–15279. https://doi.org/10.1073/pnas.2435085100

Gierahn TM, Wadsworth MH 2nd, Hughes TK, Bryson BD, Butler A, Satija R, Fortune S, Love JC, Shalek AK (2017) Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat Methods 14:395–398. https://doi.org/10.1038/nmeth.4179

Gillespie JH (1994) The causes of molecular evolution. Oxford University Press, Oxford

Glover JD, Taylor L, Sherman A, Zeiger-Poli C, Sang HM, McGrew MJ (2013) A novel piggyBac transposon inducible expression system identifies a role for AKT signalling in primordial germ cell migration. PLoS ONE 8:e77222. https://doi.org/10.1371/journal.pone.0077222

Gluckman PD, Beedle A, Hanson MA (2009) Principles of evolutionary medicine. Oxford University Press, New York

Goebel T (1999) Pleistocene human colonization of Siberia and peopling of the Americas: an ecological approach. Evol Anthropol 8:208–227

Goebel T, Waters MR, O'Rourke DH (2008) The late Pleistocene dispersal of modern humans in the Americas. Science 319:1497–1502. https://doi.org/10.1126/science.1153569

Greenberg CR, Dilling LA, Thompson GR, Seargeant LE, Haworth JC, Phillips S, Chan A, Vallance HD, Waters PJ, Sinclair G, Lillquist Y, Wanders RJ, Olpin SE (2009) The paradox of the carnitine palmitoyltransferase type Ia P479L variant in Canadian Aboriginal populations. Mol Genet Metab 96:201–207. https://doi.org/10.1016/j.ymgme.2008.12.018

Gregg XT, Prchal JT (2018) Red blood cell enzymopathies. Hematology. Elsevier, Amsterdam, pp 616–625

Grosselin K, Durand A, Marsolier J, Poitou A, Marangoni E, Nemati F, Dahmani A, Lameiras S, Reyal F, Frenoy O, Pousse Y, Reichen M, Woolfe A, Brenan C, Griffiths AD, Vallot C, Gérard A (2019) High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. Nat Genet 51:1060–1066. https://doi.org/10.1038/s41588-019-0424-9

Grossman SR, Shlyakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O, Lander ES, Schaffner SF, Sabeti PC (2010) A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327:883–886. https://doi.org/10.1126/science.1183863

Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, Park DJ, Griesemer D, Karlsson EK, Wong SH, Cabili M, Adegbola RA, Bamezai RN, Hill AV, Vannberg FO, Rinn JL, Genomes P, Lander ES, Schaffner SF, Sabeti PC (2013) Identifying recent adaptations in large-scale genomic data. Cell 152:703–713. https://doi.org/10.1016/j.cell.2013.01.035

Guenther CA, Tasic B, Luo L, Bedell MA, Kingsley DM (2014) A molecular basis for classic blond hair color in Europeans. Nat Genet 46:748–752. https://doi.org/10.1038/ng.2991

Gunther T, Nettelblad C (2019) The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLoS Genet 15:e1008302. https://doi.org/10.1371/journal.pgen.1008302

Guo J, Yang J, Visscher PM (2018) Leveraging GWAS for complex traits to detect signatures of natural selection in humans. Curr Opin Genet Dev 53:9–14. https://doi.org/10.1016/j.gde.2018.05.012

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, Fu Q,

Mittnik A, Banffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Rojo Guerra MA, Roth C, Szecsenyi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D, Cooper A, Alt KW, Reich D (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522:207–211. https://doi.org/10.1038/nature14317

Haas R, Stefanescu IC, Garcia-Putnam A, Aldenderfer MS, Clementz MT, Murphy MS, Llave CV, Watson JT (2017) Humans permanently occupied the Andean highlands by at least 7 ka. R Soc Open Sci 4:170331. https://doi.org/10.1098/rsos.170331

Haasl RJ, Payseur BA (2016) Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. Mol Ecol 25:5–23. https://doi.org/10.1111/mec.13339

Hadley TJ, Peiper SC (1997) From malaria to chemokine receptor: the emerging physiologic role of the Duffy blood group antigen. Blood 89:3077–3091

Haghverdi L, Lun ATL, Morgan MD, Marioni JC (2018) Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 36:421–427. https://doi.org/10.1038/nbt.4091

Haldane JBS (1932) The causes of evolution. Longmans Green and Co., London

Haldane JBS (1949) Disease and evolution. Supplement to La Ricerca Scientifica 19:68–76

Hamblin MT, Di Rienzo A (2000) Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am J Hum Genet 66:1669–1679. https://doi.org/10.1086/302879

Hamblin MT, Thompson EE, Di Rienzo A (2002) Complex signatures of natural selection at the Duffy blood group locus. Am J Hum Genet 70:369–383. https://doi.org/10.1086/338628

Harding RM, Healy E, Ray AJ, Ellis NS, Flanagan N, Todd C, Dixon C, Sajantila A, Jackson IJ, Birch-Machin MA, Rees JL (2000) Evidence for variable selective pressures at MC1R. Am J Hum Genet 66:1351–1361

Hardy S, Legagneux V, Audic Y, Paillard L (2010) Reverse genetics in eukaryotes. Biol Cell 102:561–580. https://doi.org/10.1042/BC20100038

Harris K (2018) The randomness that shapes our DNA. Elife. https://doi.org/10.7554/eLife.41491

Hedlund E, Deng Q (2018) Single-cell RNA sequencing: Technical advancements and biological applications. Mol Aspects Med 59:36–46. https://doi.org/10.1016/j.mam.2017.07.003

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335–2352. https://doi.org/10.1534/genetics.104.036947

Hodgson JA, Pickrell JK, Pearson LN, Quillen EE, Prista A, Rocha J, Soodyall H, Shriver MD, Perry GH (2014) Natural selection for the Duffy-null allele in the recently admixed people of Madagascar. Proc Biol Sci 281:20140930. https://doi.org/10.1098/rspb.2014.0930

Hormozdiari F, van de Bunt M, Segrè AV, Li X, Joo JWJ, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E (2016) Colocalization of GWAS and eQTL signals detects target genes. Am J Hum Genet 99:1245–1260. https://doi.org/10.1016/j.ajhg.2016.10.003

Hsieh P, Hallmark B, Watkins J, Karafet TM, Osipova LP, Gutenkunst RN, Hammer MF (2017) Exome Sequencing Provides Evidence of Polygenic Adaptation to a Fat-Rich Animal Diet in Indigenous Siberian Populations. Mol Biol Evol 34:2913–2926. https://doi.org/10.1093/molbev/msx226

Huch M, Koo BK (2015) Modeling mouse and human development using organoid cultures. Development 142:3113–3125. https://doi.org/10.1242/dev.118570

Hughes AL, Yeager M (1998) Natural selection and the evolutionary history of major histocompatibility complex loci. Front Biosci 3:d509–d516. https://doi.org/10.2741/a298

Ilardo MA, Moltke I, Korneliussen TS, Cheng J, Stern AJ, Racimo F, de Barros DP, Sikora M, Seguin-Orlando A, Rasmussen S, van den Munckhof ICL, Ter Horst R, Joosten LAB, Netea MG, Salingkat S, Nielsen R, Willerslev E (2018) Physiological and genetic adaptations to diving in sea nomads. Cell 173(569–580):e15. https://doi.org/10.1016/j.cell.2018.03.054

Ip E, Chapman G, Winlaw D, Dunwoodie SL, Giannoulatou E (2019) VPOT: a customizable variant prioritization ordering tool for annotated variants. Genom Proteom Bioinform 17:540–545. https://doi.org/10.1016/j.gpb.2019.11.001

Iskow RC, Gokcumen O, Lee C (2012) Exploring the role of copy number variants in human adaptation. Trends Genet 28:245–257. https://doi.org/10.1016/j.tig.2012.03.002

Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lonnerberg P, Linnarsson S (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods 11:163–166. https://doi.org/10.1038/nmeth.2772

Izagirre N, Garcia I, Junquera C, de la Rua C, Alonso S (2006) A scan for signatures of positive selection in candidate loci for skin pigmentation in humans. Mol Biol Evol 23:1697–1706. https://doi.org/10.1093/molbev/msl030

Jablonski NG, Chaplin G (2000) The evolution of human skin coloration. J Hum Evol 39:57–106. https://doi.org/10.1006/jhev.2000.0403

Jablonski NG, Chaplin G (2013) Epidermal pigmentation in the human lineage is an adaptation to ultraviolet radiation. J Hum Evol 65:671–675. https://doi.org/10.1016/j.jhevol.2013.06.004

Jackson L, Kuhlman C, Jackson F, Fox KP (2019) Including vulnerable populations in the assessment of data from vulnerable populations. Frontn Big Data 2:19. https://doi.org/10.3389/fdata.2019.00019

Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, Amit I (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 343:776–779. https://doi.org/10.1126/science.1247651

Jarolim P, Palek J, Amato D, Hassan K, Sapak P, Nurse GT, Rubin HL, Zhai S, Sahr KE, Liu SC (1991) Deletion in erythrocyte band 3 gene in malaria-resistant Southeast Asian ovalocytosis. Proc Natl Acad Sci USA 88:11022–11026

Jarvis JP, Scheinfeldt LB, Soi S, Lambert C, Omberg L, Ferwerda B, Froment A, Bodo JM, Beggs W, Hoffman G, Mezey J, Tishkoff SA (2012) Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. PLoS Genet 8:e1002641. https://doi.org/10.1371/journal.pgen.1002641

Jha P, Lu D, Xu S (2015) Natural selection and functional potentials of human noncoding elements revealed by analysis of next generation sequencing data. PLoS ONE 10:e0129023. https://doi.org/10.1371/journal.pone.0129023

John PR, Makova K, Li WH, Jenkins T, Ramsay M (2003) DNA polymorphism and selection at the melanocortin-1 receptor gene in normally pigmented southern African individuals. Ann N Y Acad Sci 994:299–306

Johnsson M (2018) Integrating selection mapping with genetic mapping and functional genomics. Front Genet 9:603. https://doi.org/10.3389/fgene.2018.00603

Kaidbey KH, Agin PP, Sayre RM, Kligman AM (1979) Photoprotection by melanin–a comparison of black and Caucasian skin. J Am Acad Dermatol 1:249–260

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure

in model organism association mapping. Genetics 178:1709–1723. https://doi.org/10.1534/genetics.107.080101

Kanton S, Boyle MJ, He Z, Santel M, Weigert A, Sanchís-Calleja F, Guijarro P, Sidow L, Fleck JS, Han D, Qian Z, Heide M, Huttner WB, Khaitovich P, Pääbo S, Treutlein B, Camp JG (2019) Organoid single-cell genomic atlas uncovers human-specific features of brain development. Nature 574:418–422. https://doi.org/10.1038/s41586-019-1654-9

Karczewski KJ, Snyder MP (2018) Integrative omics for health and disease. Nat Rev Genet 19:299–310. https://doi.org/10.1038/nrg.2018.4

Karlsson EK, Harris JB, Tabrizi S, Rahman A, Shlyakhter I, Patterson N, O'Dushlaine C, Schaffner SF, Gupta S, Chowdhury F, Sheikh A, Shin OS, Ellis C, Becker CE, Stuart LM, Calderwood SB, Ryan ET, Qadri F, Sabeti PC, Larocque RC (2013) Natural selection in a bangladeshi population from the cholera-endemic ganges river delta. Sci Transl Med 5:192ra86. https://doi.org/10.1126/scitranslmed.3006338

Karlsson EK, Kwiatkowski DP, Sabeti PC (2014) Natural selection and infectious disease in human populations. Nat Rev Genet 15:379–393. https://doi.org/10.1038/nrg3734

Ke M, Shen H, Wang L, Luo S, Lin L, Yang J, Tian R (2016) Identification, quantification, and site localization of protein posttranslational modifications via mass spectrometry-based proteomics. Adv Exp Med Biol 919:345–382. https://doi.org/10.1007/978-3-319-41448-5_17

Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160:765–777

Kimple ME, Brill AL, Pasker RL (2013) Overview of affinity tags for protein purification. Curr Protoc Protein Sci 73:1–23. https://doi.org/10.1002/0471140864.ps0909s73

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626. https://doi.org/10.1038/217624a0

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge Cambridgeshire, New York

Knapp M, Hofreiter M (2010) Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives. Genes (Basel) 1:227–243. https://doi.org/10.3390/genes1020227

Ko WY, Kaercher KA, Giombini E, Marcatili P, Froment A, Ibrahim M, Lema G, Nyambo TB, Omar SA, Wambebe C, Ranciaro A, Hirbo JB, Tishkoff SA (2011) Effects of natural selection and gene conversion on the evolution of human glycophorins coding for MNS blood polymorphisms in malaria-endemic African populations. Am J Hum Genet 88:741–754. https://doi.org/10.1016/j.ajhg.2011.05.005

Ko A, Cantor RM, Weissglas-Volkov D, Nikkola E, Reddy PM, Sinsheimer JS, Pasaniuc B, Brown R, Alvarez M, Rodriguez A, Rodriguez-Guillen R, Bautista IC, Arellano-Campos O, Munoz-Hernandez LL, Salomaa V, Kaprio J, Jula A, Jauhiainen M, Heliovaara M, Raitakari O, Lehtimaki T, Eriksson JG, Perola M, Lohmueller KE, Matikainen N, Taskinen MR, Rodriguez-Torres M, Riba L, Tusie-Luna T, Aguilar-Salinas CA, Pajukanta P (2014) Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. Nat Commun 5:3983. https://doi.org/10.1038/ncomms4983

Kronenberg ZN, Fiddes IT, Gordon D, Murali S, Cantsilieris S, Meyerson OS, Underwood JG, Nelson BJ, Chaisson MJP, Dougherty ML, Munson KM, Hastie AR, Diekhans M, Hormozdiari F, Lorusso N, Hoekzema K, Qiu R, Clark K, Raja A, Welch AE, Sorensen M, Baker C, Fulton RS, Armstrong J, Graves-Lindsay TA, Denli AM, Hoppe ER, Hsieh P, Hill CM, Pang AWC, Lee J, Lam ET, Dutcher SK, Gage FH, Warren WC, Shendure J, Haussler D, Schneider VA, Cao H, Ventura M, Wilson RK, Paten B, Pollen A, Eichler EE (2018) High-resolution comparative analysis of great ape genomes. Science. https://doi.org/10.1126/science.aar6343

Kwiatkowski DP (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. Am J Hum Genet 77:171–192. https://doi.org/10.1086/432519

Laland KN, Odling-Smee J, Myles S (2010) How culture shaped the human genome: bringing genetics and the human sciences together. Nat Rev Genet 11:137–148. https://doi.org/10.1038/nrg2734

Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, Sinha S, Moore JL, Jagadeeswaran P, Zhao W, Ning G, Makalowska I, McKeigue PM, O'Donnell D, Kittles R, Parra EJ, Mangini NJ, Grunwald DJ, Shriver MD, Canfield VA, Cheng KC (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science 310:1782–1786. https://doi.org/10.1126/science.1116238

Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME, Homfray T, Penninger JM, Jackson AP, Knoblich JA (2013) Cerebral organoids model human brain development and microcephaly. Nature 501:373–379. https://doi.org/10.1038/nature12517

Langer-Safer PR, Levine M, Ward DC (1982) Immunological method for mapping genes on *Drosophila* polytene chromosomes. Proc Natl Acad Sci USA 79:4381–4385. https://doi.org/10.1073/pnas.79.14.4381

Lao O, de Gruijter JM, van Duijn K, Navarro A, Kayser M (2007) Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. Ann Hum Genet 71:354–369. https://doi.org/10.1111/j.1469-1809.2006.00341.x

Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. Am Nat 151:283–292. https://doi.org/10.1086/286119

Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait land trait in a subdivided population under selection. Genetics 164:1205–1219

Lee S, Wang H, Xing EP (2017) Backward genotype-transcript-phenotype association mapping. Methods 129:18–23. https://doi.org/10.1016/j.ymeth.2017.09.004

Leffler EM, Band G, Busby GBJ, Kivinen K, Le QS, Clarke GM, Bojang KA, Conway DJ, Jallow M, Sisay-Joof F, Bougouma EC, Mangano VD, Modiano D, Sirima SB, Achidi E, Apinjoh TO, Marsh K, Ndila CM, Peshu N, Williams TN, Drakeley C, Manjurano A, Reyburn H, Riley E, Kachala D, Molyneux M, Nyirongo V, Taylor T, Thornton N, Tilley L, Grimsley S, Drury E, Stalker J, Cornelius V, Hubbart C, Jeffreys AE, Rowlands K, Rockett KA, Spencer CCA, Kwiatkowski DP, Network MGE (2017) Resistance to malaria through structural variation of red blood cell invasion receptors. Science. https://doi.org/10.1126/science.aam6393

Lemas DJ, Wiener HW, O'Brien DM, Hopkins S, Stanhope KL, Havel PJ, Allison DB, Fernandez JR, Tiwari HK, Boyer BB (2012) Genetic polymorphisms in carnitine palmitoyltransferase 1A gene are associated with variation in body composition and fasting lipid traits in Yup'ik Eskimos. J Lipid Res 53:175–184. https://doi.org/10.1194/jlr.P018952

Leonelli S, Ankeny RA (2013) What makes a model organism? Endeavour 37:209–212

León-Velarde F (2003) Pursuing international recognition of chronic mountain sickness. High Alt Med Biol 4:256–259. https://doi.org/10.1089/152702903322022857

Lewandowska M, Jedrychowska-Danska K, Ploszaj T, Witas P, Zamerska A, Mankowska-Pliszka H, Witas HW (2018) Searching for signals of recent natural selection in genes of the innate immune response: ancient DNA study. Infect Genet Evol 63:62–72. https://doi.org/10.1016/j.meegid.2018.05.008

Li Y, Muffat J, Omer A, Bosch I, Lancaster MA, Sur M, Gehrke L, Knoblich JA, Jaenisch R (2017) Induction of expansion and folding in human cerebral organoids. Cell Stem Cell 20:385–396.e3. https://doi.org/10.1016/j.stem.2016.11.017

Libert FCP, Beckman G, Samson M, Aksenova M et al (1998) The delta CCR5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. Hum Mol Genet 7:399–406

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293. https://doi.org/10.1126/science.1181369

Lim JK, Lisco A, McDermott DH, Huynh L, Ward JM, Johnson B, Johnson H, Pape J, Foster GA, Krysztof D, Follmann D, Stramer SL, Margolis LB, Murphy PM (2009) Genetic variation in OAS1 is a risk factor for initial infection with West Nile virus in man. PLoS Pathog 5:e1000321. https://doi.org/10.1371/journal.ppat.1000321

Lindo J, Haas R, Hofman C, Apata M, Moraga M, Verdugo RA, Watson JT, Viviano Llave C, Witonsky D, Beall C, Warinner C, Novembre J, Aldenderfer M, Di Rienzo A (2018) The genetic prehistory of the Andean highlands 7000 years BP though European contact. Sci Adv 4:eaau4921. https://doi.org/10.1126/sciadv.aau4921

Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, MacDonald ME, Stuhlmann H, Koup RA, Landau NR (1996) Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 86:367–377

Liu X, Yunus Y, Lu D, Aghakhanian F, Saw WY, Deng L, Ali M, Wang X, Nor FM, Rahman TA, Shaari SA, Salleh MZ, Phipps ME, Ong RT, Xu S, Teo YY, Hoh BP (2015) Differential positive selection of malaria resistance genes in three indigenous populations of Peninsular Malaysia. Hum Genet 134:375–392. https://doi.org/10.1007/s00439-014-1525-2

Livingstone FB (1984) The Duffy blood groups, vivax malaria, and malaria selection in human populations: a review. Hum Biol 56:413–425

Loomis WF (1967) Skin-pigment regulation of vitamin-D biosynthesis in man. Science 157:501–506

Lorenzo FR, Huff C, Myllymäki M, Olenchock B, Swierczek S, Tashi T, Gordeuk V, Wuren T, Ri-Li G, McClain DA, Khan TM, Koul PA, Guchhait P, Salama ME, Xing J, Semenza GL, Liberzon E, Wilson A, Simonson TS, Jorde LB, Kaelin WG, Koivunen P, Prchal JT (2014) A genetic mechanism for Tibetan high-altitude adaptation. Nat Genet 46:951–956. https://doi.org/10.1038/ng.3067

Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T (2017) Transcriptomics technologies. PLoS Comput Biol 13:e1005457. https://doi.org/10.1371/journal.pcbi.1005457

Lucas ER, Miles A, Harding NJ, Clarkson CS, Lawniczak MKN, Kwiatkowski DP, Weetman D, Donnelly MJ, Consortium AgG (2019) Whole-genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. Genome Res 29:1250–1261. https://doi.org/10.1101/gr.245795.118

Luzzatto L (2012) Sickle cell anaemia and malaria. Mediterr J Hematol Infect Dis 4:e2012065. https://doi.org/10.4084/MJHID.2012.065

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161:1202–1214. https://doi.org/10.1016/j.cell.2015.05.002

Maddocks S, Jenkins R (2017) Chapter 4- quantitative PCR: things to consider. understanding pCR: a practical bench-top guide. Academic Press, London, pp 45–52

Mahmood T, Yang PC (2012) Western blot: technique, theory, and trouble shooting. N Am J Med Sci 4:429–434. https://doi.org/10.4103/1947-2714.100998

Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, Ferrari R (2018) Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief Bioinform 19:286–302. https://doi.org/10.1093/bib/bbw114

Marchetti G, Pinotti M, Lunghi B, Casari C, Bernardi F (2012) Functional genetics. Thromb Res 129:336–340. https://doi.org/10.1016/j.thromres.2011.10.028

Marth GT, Czabarka E, Murvai J, Sherry ST (2004) The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics 166:351–372

Martins-de-Souza D (2014) Proteomics, metabolomics, and protein interactomics in the characterization of the molecular features of major depressive disorder. Dialog Clin Neurosci 16:63–73

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, de Castro JM, Carbonell E, Gerritsen F, Khokhlov A, Kuznetsov P, Lozano M, Meller H, Mochalov O, Moiseyev V, Guerra MA, Roodenberg J, Verges JM, Krause J, Cooper A, Alt KW, Brown D, Anthony D, Lalueza-Fox C, Haak W, Pinhasi R, Reich D (2015) Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528:499–503. https://doi.org/10.1038/nature16152

Matsumoto Y, Goto T, Nishino J, Nakaoka H, Tanave A, Takano-Shimizu T, Mott RF, Koide T (2017) Selective breeding and selection mapping using a novel wild-derived heterogeneous stock of mice revealed two closely-linked loci for tameness. Sci Rep 7:4607. https://doi.org/10.1038/s41598-017-04869-1

Mayer S, Chen J, Velmeshev D, Mayer A, Eze UC, Bhaduri A, Cunha CE, Jung D, Arjun A, Li E, Alvarado B, Wang S, Lovegren N, Gonzales ML, Szpankowski L, Leyrat A, West JAA, Panagiotakos G, Alvarez-Buylla A, Paredes MF, Nowakowski TJ, Pollen AA, Kriegstein AR (2019) Multimodal single-cell analysis reveals physiological maturation in the developing human neocortex. Neuron 102:143–158.e7. https://doi.org/10.1016/j.neuron.2019.01.027

Maynard-Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23:23–35

McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X, Menke DB, Schaar BT, Wenger AM, Bejerano G, Kingsley DM (2011) Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature 471:216–219. https://doi.org/10.1038/nature09774

McVicker G, Gordon D, Davis C, Green P (2009) Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet 5:e1000471. https://doi.org/10.1371/journal.pgen.1000471

Meyer CG, Calixto Fernandes MH, Intemann CD, Kreuels B, Kobbe R, Kreuzberg C, Ayim M, Ruether A, Loag W, Ehmen C, Adjei S, Adjei O, Horstmann RD, May J (2011) IL3 variant on chromosomal region 5q31-33 and protection from recurrent malaria attacks. Hum Mol Genet 20:1173–1181. https://doi.org/10.1093/hmg/ddq562

Miller LH, Mason SJ, Clyde DF, McGinniss MH (1976) The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype. FyFy N Engl J Med 295:302–304. https://doi.org/10.1056/NEJM197608052950602

Monge C (1942) Life in the andes and chronic mountain sickness. Science 95:79–84. https://doi.org/10.1126/science.95.2456.79
Monroy Kuhn JM, Jakobsson M, Gunther T (2018) Estimating genetic kin relationships in prehistoric populations. PLoS ONE 13:e0195491. https://doi.org/10.1371/journal.pone.0195491
Musunuru K, Bernstein D, Cole FS, Khokha MK, Lee FS, Lin S, McDonald TV, Moskowitz IP, Quertermous T, Sankaran VG, Schwartz DA, Silverman EK, Zhou X, Hasan AAK, Luo XJ (2018) Functional assays to screen and dissect genomic hits: doubling down on the National Investment in Genomic Research. Circ Genom Precis Med 11:e002178. https://doi.org/10.1161/CIRCGEN.118.002178
Myles S, Somel M, Tang K, Kelso J, Stoneking M (2007) Identifying genes underlying skin pigmentation differences among human populations. Hum Genet 120:613–621. https://doi.org/10.1007/s00439-006-0256-4
Nesse RM, Stearns SC (2008) The great opportunity: Evolutionary applications to medicine and public health. Evol Appl 1:28–48. https://doi.org/10.1111/j.1752-4571.2007.00006.x
Nesse RM, Bergstrom CT, Ellison PT, Flier JS, Gluckman P, Govindaraju DR, Niethammer D, Omenn GS, Perlman RL, Schwartz MD, Thomas MG, Stearns SC, Valle D (2010) Evolution in health and medicine Sackler colloquium: making evolutionary biology a basic science for medicine. Proc Natl Acad Sci USA 107(Suppl 1):1800–1807. https://doi.org/10.1073/pnas.0906224106
Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for selective sweeps using SNP data. Genome Res 15:1566–1575. https://doi.org/10.1101/gr.4252305
Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG (2007) Recent and ongoing selection in the human genome. Nat Rev Genet 8:857–868. https://doi.org/10.1038/nrg2187
Niermeyer S, Zamdio S, Moore LG (2001) The people. In: Hornbein TF, Schoene RB (eds) High altitude: an exploration of human adaptation. Marcel Dekker, New York, pp 42–100
Nish S, Medzhitov R (2011) Host defense pathways: role of redundancy and compensation in infectious disease phenotypes. Immunity 34:629–636. https://doi.org/10.1016/j.immuni.2011.05.009
Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18:83–90. https://doi.org/10.1016/s0168-9525(02)02557-x
Norton HL, Kittles RA, Parra E, McKeigue P, Mao X, Cheng K, Canfield VA, Bradley DG, McEvoy B, Shriver MD (2007) Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Mol Biol Evol 24:710–722. https://doi.org/10.1093/molbev/msl203
Novembre J, Galvani AP, Slatkin M (2005) The geographic spread of the CCR5 Delta32 HIV-resistance allele. PLoS Biol 3:e339. https://doi.org/10.1371/journal.pbio.0030339
Oliveira TY, Harris EE, Meyer D, Jue CK, Silva WA (2012) Molecular evolution of a malaria resistance gene (DARC) in primates. Immunogenetics 64:497–505. https://doi.org/10.1007/s00251-012-0608-2
Paquette AM, Harahap A, Laosombat V, Patnode JM, Satyagraha A, Sudoyo H, Thompson MK, Yusoff NM, Wilder JA (2015) The evolutionary origins of Southeast Asian Ovalocytosis. Infect Genet Evol 34:153–159. https://doi.org/10.1016/j.meegid.2015.06.002
Park IH, Lerou PH, Zhao R, Huo H, Daley GQ (2008) Generation of human-induced pluripotent stem cells. Nat Protoc 3:1180–1186
Peirson SN, Butler JN (2007) Quantitative polymerase chain reaction. Methods Mol Biol 362:349–362. https://doi.org/10.1007/978-1-59745-257-1_25
Peng X, Emiliani F, Smallwood PM, Rattner A, Lei H, Sabbagh MF, Nathans J (2018) Affinity capture of polyribosomes followed by RNAseq (ACAPseq), a discovery platform for protein-protein interactions. Elife. https://doi.org/10.7554/eLife.40982
Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC (2007) Diet and the evolution of human amylase gene copy number variation. Nat Genet 39:1256–1260. https://doi.org/10.1038/ng2123
Peter BM, Huerta-Sanchez E, Nielsen R (2012) Distinguishing between selective sweeps from standing variation and from a de novo mutation. PLoS Genet 8:e1003011. https://doi.org/10.1371/journal.pgen.1003011
Phizicky EM, Fields S (1995) Protein-protein interactions: methods for detection and analysis. Microbiol Rev 59:94–123
Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, Pritchard JK (2009) Signals of recent positive selection in a worldwide sample of human populations. Genome Res 19:826–837. https://doi.org/10.1101/gr.087577.108
Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, Williams TN, Weatherall DJ, Hay SI (2010) Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. Nat Commun 1:104. https://doi.org/10.1038/ncomms1104
Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS (2018) Long reads: their purpose and place. Hum Mol Genet 27:R234–R241. https://doi.org/10.1093/hmg/ddy177
Pollen AA, Bhaduri A, Andrews MG, Nowakowski TJ, Meyerson OS, Mostajo-Radji MA, Di Lullo E, Alvarado B, Bedolli M, Dougherty ML, Fiddes IT, Kronenberg ZN, Shuga J, Leyrat AA, West JA, Bershteyn M, Lowe CB, Pavlovic BJ, Salama SR, Haussler D, Eichler EE, Kriegstein AR (2019) Establishing cerebral organoids as models of human-specific brain evolution. Cell 176:743–756.e17. https://doi.org/10.1016/j.cell.2019.01.017
Popejoy AB, Fullerton SM (2016) Genomics is failing on diversity. Nature 538:161–164. https://doi.org/10.1038/538161a
Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol 20:R208–R215. https://doi.org/10.1016/j.cub.2009.11.055
Przeworski M (2002) The signature of positive selection at randomly chosen loci. Genetics 160:1179–1189
Przeworski M, Hudson RR, Di Rienzo A (2000) Adjusting the focus on human variation. Trends Genet 16:296–302. https://doi.org/10.1016/s0168-9525(00)02030-8
Przeworski M, Coop G, Wall JD (2005) The signature of positive selection on standing genetic variation. Evolution 59:2312–2323
Quillen EE, Bauchet M, Bigham AW, Delgado-Burbano ME, Faust FX, Klimentidis YC, Mao X, Stoneking M, Shriver MD (2012) OPRM1 and EGFR contribute to skin pigmentation differences between Indigenous Americans and Europeans. Hum Genet 131:1073–1080. https://doi.org/10.1007/s00439-011-1135-1
Racimo F (2016) Testing for ancient selection using cross-population allele frequency differentiation. Genetics 202:733. https://doi.org/10.1534/genetics.115.178095
Raha D, Hong M, Snyder M (2010) ChIP-Seq: a method for global identification of regulatory elements in the genome. Curr Protoc Mol Biol. https://doi.org/10.1002/0471142727.mb2119s91
Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8:2281–2308. https://doi.org/10.1038/nprot.2013.143
Rana BK, Hewett-Emmett D, Jin L, Chang BH, Sambuughin N, Lin M, Watkins S, Bamshad MB, Jorde LB, Ramsay M, Jenkins T, Li WH (1999) High polymorphism at the human melanocortin 1 receptor locus. Genetics 151:1547–1557

Reed LK, Baer CF, Edison AS (2017) Considerations when choosing a genetic model organism for metabolomics studies. Curr Opin Chem Biol 36:7–14. https://doi.org/10.1016/j.cbpa.2016.12.005

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517. https://doi.org/10.1126/science.273.5281.1516

Rodriguez-Murillo L, Greenberg DA (2008) Genetic association analysis: a primer on how it works, its strengths and its weaknesses. Int J Androl 31:546–556. https://doi.org/10.1111/j.1365-2605.2008.00896.x

Rumold CU, Aldenderfer MS (2016) Late Archaic-Early Formative period microbotanical evidence for potato at Jiskairumoko in the Titicaca Basin of southern Peru. Proc Natl Acad Sci USA 113:13672–13677. https://doi.org/10.1073/pnas.1604265113

Ruwende C, Khoo SC, Snow RW, Yates SN, Kwiatkowski D, Gupta S, Warn P, Allsopp CE, Gilbert SC, Peschu N (1995) Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. Nature 376:246–249. https://doi.org/10.1038/376246a0

Sabeti P, Usen S, Farhadian S, Jallow M, Doherty T, Newport M, Pinder M, Ward R, Kwiatkowski D (2002a) CD40L association with protection from severe malaria. Genes Immun 3:286–291. https://doi.org/10.1038/sj.gene.6363877

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002b) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837. https://doi.org/10.1038/nature01140

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD et al (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449:913–918. https://doi.org/10.1038/nature06250

Salas D, Stacey RG, Akinlaja M, Foster LJ (2020) Next-generation interactomics: considerations for the use of co-elution to measure protein interaction networks. Mol Cell Proteom 19:1–10. https://doi.org/10.1074/mcp.R119.001803

Sambo MR, Trovoada MJ, Benchimol C, Quinhentos V, Gonçalves L, Velosa R, Marques MI, Sepúlveda N, Clark TG, Mustafa S, Wagner O, Coutinho A, Penha-Gonçalves C (2010) Transforming growth factor beta 2 and heme oxygenase 1 genes are risk factors for the cerebral malaria syndrome in Angolan children. PLoS ONE 5:e11141. https://doi.org/10.1371/journal.pone.0011141

Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, Saragosti S, Lapoumeroulie C, Cognaux J, Forceille C, Muyldermans G, Verhofstede C, Burtonboy G, Georges M, Imai T, Rana S, Yi Y, Smyth RJ, Collman RG, Doms RW, Vassart G, Parmentier M (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382:722–725. https://doi.org/10.1038/382722a0

Sapkota D, Lake AM, Yang W, Yang C, Wesseling H, Guise A, Uncu C, Dalal JS, Kraft AW, Lee JM, Sands MS, Steen JA, Dougherty JD (2019) Cell-type-specific profiling of alternative translation identifies regulated protein isoform variation in the mouse brain. Cell Rep 26:594–607.e7. https://doi.org/10.1016/j.celrep.2018.12.077

Saunders MA, Hammer MF, Nachman MW (2002) Nucleotide variability at G6pd and the signature of malarial selection in humans. Genetics 162:1849–1861

Saunders MA, Slatkin M, Garner C, Hammer MF, Nachman MW (2005) The extent of linkage disequilibrium caused by selection on G6PD in humans. Genetics 171:1219–1229. https://doi.org/10.1534/genetics.105.048140

Schaid DJ, Chen W, Larson NB (2018) From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet 19:491–504. https://doi.org/10.1038/s41576-018-0016-z

Scheinfeldt LB, Tishkoff SA (2013) Recent human adaptation: genomic approaches, interpretation and insights. Nat Rev Genet 14:692–702. https://doi.org/10.1038/nrg3604

Scheinfeldt LB, Soi S, Thompson S, Ranciaro A, Woldemeskel D, Beggs W, Lambert C, Jarvis JP, Abate D, Belay G, Tishkoff SA (2012) Genetic adaptation to high altitude in the Ethiopian highlands. Genome Biol 13:R1. https://doi.org/10.1186/gb-2012-13-1-r1

Schlebusch CM, Lewis CM Jr, Vahter M, Engstrom K, Tito RY, Obregon-Tito AJ, Huerta D, Polo SI, Medina AC, Brutsaert TD, Concha G, Jakobsson M, Broberg K (2013) Possible positive selection for an arsenic-protective haplotype in humans. Environ Health Perspect 121:53–58. https://doi.org/10.1289/ehp.1205504

Schmitz S, Thomas PD, Allen TM, Poznansky MJ, Jimbow K (1995) Dual role of melanins and melanin precursors as photoprotective and phototoxic agents: inhibition of ultraviolet radiation-induced lipid peroxidation. Photochem Photobiol 61:650–655

Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH (2001) The origins of acquired immune deficiency syndrome viruses: where and when? Philos Trans R Soc Lond B Biol Sci 356:867–876. https://doi.org/10.1098/rstb.2001.0863

Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, Jones KW (2004) The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Hum Genom 1:274–286

Sim BK, Chitnis CE, Wasniowska K, Hadley TJ, Miller LH (1994) Receptor and ligand domains for invasion of erythrocytes by Plasmodium falciparum. Science 264:1941–1944. https://doi.org/10.1126/science.8009226

Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, Prchal JT, Ge R (2010) Genetic evidence for high-altitude adaptation in Tibet. Science 329:72–75. https://doi.org/10.1126/science.1189406

Simonson TS, McClain DA, Jorde LB, Prchal JT (2012) Genetic determinants of Tibetan high-altitude adaptation. Hum Genet 131:527–533. https://doi.org/10.1007/s00439-011-1109-3

Singh BD, Singh AK (2015) Association mapping. Marker-assisted plant breeding: principles and practices. Springer, New Delhi, pp 217–256

Sirugo G, Williams SM, Tishkoff SA (2019) The missing diversity in human genetic studies. Cell 177:1080. https://doi.org/10.1016/j.cell.2019.04.032

Smemo S, Campos LC, Moskowitz IP, Krieger JE, Pereira AC, Nobrega MA (2012) Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. Hum Mol Genet 21:3255–3263. https://doi.org/10.1093/hmg/dds165

Song D, Navalsky BE, Guan W, Ingersoll C, Wang T, Loro E, Eeles L, Matchett KB, Percy MJ, Walsby-Tickle J, McCullagh JSO, Medina RJ, Khurana TS, Bigham AW, Lappin TR, Lee FS (2020)

Tibetan PHD2, an allele with loss-of-function properties. PNAS 117:12230–12238
Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, Winkler C, Huttley GA, Allikmets R, Schriml L, Gerrard B, Malasky M, Ramos MD, Morlot S, Tzetis M, Oddoux C, di Giovine FS, Nasioulas G, Chandler D, Aseev M, Hanson M, Kalaydjieva L, Glavac D, Gasparini P, Kanavakis E, Claustres M, Kambouris M, Ostrer H, Duff G, Baranov V, Sibul H, Metspalu A, Goldman D, Martin N, Duffy D, Schmidtke J, Estivill X, O'Brien SJ, Dean M (1998) Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. Am J Hum Genet 62:1507–1515. https://doi.org/10.1086/301867
Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R (2019) Comprehensive integration of single-cell data. Cell 177:1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031
Sugden LA, Atkinson EG, Fischer AP, Rong S, Henn BM, Ramachandran S (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
Suktitipat B, Naktang C, Mhuantong W, Tularak T, Artiwet P, Pasomsap E, Jongjaroenprasert W, Fuchareon S, Mahasirimongkol S, Chantratita W, Yimwadsana B, Charoensawan V, Jinawath N (2014) Copy number variation in Thai population. PLoS ONE 9:e104355. https://doi.org/10.1371/journal.pone.0104355
Supharattanasitthi W, Carlsson E, Sharif U, Paraoan L (2019) CRISPR/Cas9-mediated one step bi-allelic change of genomic DNA in iPSCs and human RPE cells in vitro with dual antibiotic selection. Sci Rep 9:174. https://doi.org/10.1038/s41598-018-36740-2
Suzuki IK, Gacquer D, Van Heurck R, Kumar D, Wojno M, Bilheu A, Herpoel A, Lambert N, Cheron J, Polleux F, Detours V, Vanderhaeghen P (2018) Human-specific NOTCH2NL genes expand cortical neurogenesis through delta/notch regulation. Cell 173(1370–1384):e16. https://doi.org/10.1016/j.cell.2018.03.067
Szabo Q, Bantignies F, Cavalli G (2019) Principles of genome folding into topologically associating domains. Sci Adv 5:eaaw1668. https://doi.org/10.1126/sciadv.aaw1668
Szpak M, Mezzavilla M, Ayub Q, Chen Y, Xue Y, Tyler-Smith C (2018) FineMAV: prioritizing candidate genetic variants driving local adaptations in human populations. Genome Biol 19:5. https://doi.org/10.1186/s13059-017-1380-2
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595
Tak YG, Farnham PJ (2015) Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenetics Chromatin 8:57. https://doi.org/10.1186/s13072-015-0050-4
Tashi T, Scott Reading N, Wuren T, Zhang X, Moore LG, Hu H, Tang F, Shestakova A, Lorenzo F, Burjanivova T, Koul P, Guchhait P, Wittwer CT, Julian CG, Shah B, Huff CD, Gordeuk VR, Prchal JT, Ge R (2017) Gain-of-function EGLN1 prolyl hydroxylase (PHD2 D4E:C127S) in combination with EPAS1 (HIF-2α) polymorphism lowers hemoglobin concentration in Tibetan highlanders. J Mol Med (Berl) 95:665–670. https://doi.org/10.1007/s00109-017-1519-3
Timme-Laragy AR, Karchner SI, Hahn ME (2012) Gene knockdown by morpholino-modified oligonucleotides in the zebrafish (Danio rerio) model: applications for developmental toxicology. Methods Mol Biol 889:51–71. https://doi.org/10.1007/978-1-61779-867-2_5
Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J, Piro A, Stoneking M, Tagarelli A, Tagarelli G, Touma EH, Williams SM, Clark AG (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science 293:455–462. https://doi.org/10.1126/science.1061573
Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P (2007) Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 39:31–40. https://doi.org/10.1038/ng1946
Tolia NH, Enemark EJ, Sim BK, Joshua-Tor L (2005) Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite Plasmodium falciparum. Cell 122:183–193. https://doi.org/10.1016/j.cell.2005.05.033
van de Bunt M, Cortes A, Brown MA, Morris AP, McCarthy MI, Consortium I (2015) Evaluating the performance of fine-mapping strategies at common variant GWAS loci. PLoS Genet 11:e1005535. https://doi.org/10.1371/journal.pgen.1005535
Verma N, Zhu Z, Huangfu D (2017) CRISPR/Cas-mediated knockin in human pluripotent stem cells. Methods Mol Biol 1513:119–140. https://doi.org/10.1007/978-1-4939-6539-7_9
Verrelli BC, McDonald JH, Argyropoulos G, Destro-Bisol G, Froment A, Drousiotou A, Lefranc G, Helal AN, Loiselet J, Tishkoff SA (2002) Evidence for balancing selection from nucleotide sequence analyses of human G6PD. Am J Hum Genet 71:1112–1128. https://doi.org/10.1086/344345
Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting natural selection in genomic data. Annu Rev Genet 47:97–120. https://doi.org/10.1146/annurev-genet-111212-133526
Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4:e72. https://doi.org/10.1371/journal.pbio.0040072
Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484
Wang M, Huang X, Li R, Xu H, Jin L, He Y (2014) Detecting recent positive selection with high accuracy and reliability by conditional coalescent tree. Mol Biol Evol 31:3068–3080. https://doi.org/10.1093/molbev/msu244
Wang J, Huang D, Zhou Y, Yao H, Liu H, Zhai S, Wu C, Zheng Z, Zhao K, Wang Z, Yi X, Zhang S, Liu X, Liu Z, Chen K, Yu Y, Sham PC, Li MJ (2020) CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies. Nucleic Acids Res 48:D807–D816. https://doi.org/10.1093/nar/gkz1026
White BA (1997) PCR cloning protocols: from molecular cloning to genetic engineering. Springer, Berlin
Wiener P, Pong-Wong R (2011) A regression-based approach to selection mapping. J Hered 102:294–305. https://doi.org/10.1093/jhered/esr014
Wilder JA, Stone JA, Preston EG, Finn LE, Ratcliffe HL, Sudoyo H (2009) Molecular population genetics of SLC4A1 and Southeast Asian ovalocytosis. J Hum Genet 54:182–187. https://doi.org/10.1038/jhg.2009.12
Williams GC, Nesse RM (1991) The dawn of Darwinian medicine. Q Rev Biol 66:1–22. https://doi.org/10.1086/417048
Wisser RJ, Murray SC, Kolkman JM, Ceballos H, Nelson RJ (2008) Selection mapping of loci for quantitative disease resistance in a diverse maize population. Genetics 180:583–599. https://doi.org/10.1534/genetics.108.090118
Wooding SP, Watkins WS, Bamshad MJ, Dunn DM, Weiss RB, Jorde LB (2002) DNA sequence variation in a 3.7-kb noncoding sequence 5' of the CYP1A2 gene: implications for human population history and natural selection. Am J Hum Genet 71:528–542. https://doi.org/10.1086/342260
Wright S (1950) Genetical structure of populations. Nature 166:247–249

Wu DD, Li GM, Jin W, Li Y, Zhang YP (2012) Positive selection on the osteoarthritis-risk and decreased-height associated variants at the GDF5 gene in East Asians. PLoS ONE 7:e42553. https://doi.org/10.1371/journal.pone.0042553

Xu Z, Wu C, Wei P, Pan W (2017) A Powerful framework for integrating eQTL and GWAS summary data. Genetics 207:893–902. https://doi.org/10.1534/genetics.117.300270

Yakub I, Lillibridge KM, Moran A, Gonzalez OY, Belmont J, Gibbs RA, Tweardy DJ (2005) Single nucleotide polymorphisms in genes for 2'-5'-oligoadenylate synthetase and RNase L inpatients hospitalized with West Nile virus infection. J Infect Dis 192:1741–1748. https://doi.org/10.1086/497340

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, Zheng H, Liu T, He W, Li K, Luo R, Nie X, Wu H, Zhao M, Cao H, Zou J, Shan Y, Li S, Yang Q, Asan NP, Tian G, Xu J, Liu X, Jiang T, Wu R, Zhou G, Tang M, Qin J, Wang T, Feng S, Li G, Huasang LJ, Wang W, Chen F, Wang Y, Zheng X, Li Z, Bianba Z, Yang G, Wang X, Tang S, Gao G, Chen Y, Luo Z, Gusang L, Cao Z, Zhang Q, Ouyang W, Ren X, Liang H, Zheng H, Huang Y, Li J, Bolund L, Kristiansen K, Li Y, Zhang Y, Zhang X, Li R, Li S, Yang H, Nielsen R, Wang J, Wang J (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329:75–78. https://doi.org/10.1126/science.1190371

Yin X, Mead BE, Safaee H, Langer R, Karp JM, Levy O (2016) Engineering stem cell organoids. Cell Stem Cell 18:25–38. https://doi.org/10.1016/j.stem.2015.12.005

Zhang C, Bailey DK, Awad T, Liu G, Xing G, Cao M, Valmeekam V, Retief J, Matsuzaki H, Taub M, Seielstad M, Kennedy GC (2006) A whole genome long-range haplotype (WGLRH) test for detecting imprints of positive selection in human populations. Bioinformatics 22:2122–2128. https://doi.org/10.1093/bioinformatics/btl365

Zhang S, Guo F, Yan W, Dai Z, Dong W, Zhou J, Zhang W, Xin F, Jiang M (2019) Recent advances of CRISPR/Cas9-based genetic engineering and transcriptional regulation in industrial biology. Front Bioeng Biotechnol 7:459. https://doi.org/10.3389/fbioe.2019.00459

Zhao Z, Jin L, Fu YX, Ramsay M, Jenkins T, Leskinen E, Pamilo P, Trexler M, Patthy L, Jorde LB, Ramos-Onsins S, Yu N, Li WH (2000) Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc Natl Acad Sci USA 97:11354–11358. https://doi.org/10.1073/pnas.200348197

Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM (2014) Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics 15:419. https://doi.org/10.1186/1471-2164-15-419

Zhou H, Ding S (2010) Evolution of induced pluripotent stem cell technology. Curr Opin Hematol 17:276–280. https://doi.org/10.1097/MOH.0b013e328339f2ee

Zhou L, Zhao F (2018) Prioritization and functional assessment of noncoding variants associated with complex diseases. Genome Med 10:53. https://doi.org/10.1186/s13073-018-0565-y