

Human genomic disease variants: A neutral evolutionary explanation

Joel T. Dudley,^{1,2,3} Yuseob Kim,^{4,5} Li Liu,⁵ Glenn J. Markov,^{5,6} Kristyn Gerold,⁵ Rong Chen,² Atul J. Butte,^{1,2,3} and Sudhir Kumar^{5,7,8}

¹Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, California 94305, USA; ²Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, California 94305, USA; ³Lucile Packard Children's Hospital, Palo Alto, California 94304, USA; ⁴Department of Life Science, Ewha Womans University, Seoul 120-750, Korea; ⁵Center for Evolutionary Medicine and Informatics, Biodesign Institute, Arizona State University, Tempe, Arizona 85287-5301, USA; ⁶Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA; ⁷School of Life Sciences, Arizona State University, Tempe, Arizona 85287-5301, USA

Many perspectives on the role of evolution in human health include nonempirical assumptions concerning the adaptive evolutionary origins of human diseases. Evolutionary analyses of the increasing wealth of clinical and population genomic data have begun to challenge these presumptions. In order to systematically evaluate such claims, the time has come to build a common framework for an empirical and intellectual unification of evolution and modern medicine. We review the emerging evidence and provide a supporting conceptual framework that establishes the classical neutral theory of molecular evolution (NTME) as the basis for evaluating disease-associated genomic variations in health and medicine. For over a decade, the NTME has already explained the origins and distribution of variants implicated in diseases and has illuminated the power of evolutionary thinking in genomic medicine. We suggest that a majority of disease variants in modern populations will have neutral evolutionary origins (previously neutral), with a relatively smaller fraction exhibiting adaptive evolutionary origins (previously adaptive). This pattern is expected to hold true for common as well as rare disease variants. Ultimately, a neutral evolutionary perspective will provide medicine with an informative and actionable framework that enables objective clinical assessment beyond convenient tendencies to invoke past adaptive events in human history as a root cause of human disease.

Since the announcements of Darwin's great discovery in the 19th century, many scientists have shown an interest in the applications of evolutionary principles into medicine (Nesse and Williams 1996; Stearns 1999; Trevathan et al. 1999; Kumar et al. 2011). This interface between evolutionary biology and medicine became more broadly apparent after the establishment of DNA as the hereditary material in the 1950s, and subsequent advances in DNA sequencing technologies over the past two decades. Scores of investigations have highlighted the potential of applying evolutionary principles in human health and disease (McKenna et al. 1993; Miller and Kumar 2001; Ramensky et al. 2002; Nesse et al. 2006; Kumar et al. 2011). These include recent evolutionary insights into the origins of specific, clinically relevant phenotypes, such as lactose tolerance and malaria resistance (Sabeti et al. 2006; Tishkoff et al. 2006; Williamson et al. 2007; Tung et al. 2009). The potential to apply evolutionary methods and principles toward clinical utility is gaining in significance with increasing public availability, and declining production costs, of sequence data measured from individuals and populations (Feero et al. 2010; Kumar et al. 2011). Consequently, efforts to incorporate evolutionary biology into medical education are now underway and expanding (Nesse et al. 2009; Stearns et al. 2010).

As the interface between evolutionary biology and genomic medicine progresses into the mainstream of clinical research

and training, it is becoming important to establish conceptual frameworks for exploring this interface based on sound null hypotheses. Such frameworks are a prerequisite for the scientific evaluation of claims concerning the role or utility of evolutionary biology in medicine (Miller and Kumar 2001; Gluckman et al. 2009; Nesse et al. 2009; Kumar et al. 2011). Here we suggest, with strong support from the empirical data, that the classical neutral theory of molecular evolution (NTME) (Kimura 1983) provides a ready, validated framework for generating a null hypothesis required for the discovery, characterization, and clinical evaluation of human disease-associated variation. For the past 50 yr, the NTME has firmly established principles that explain the fate of the majority of new mutations in a population, and the emergence of differences within and among species (see below). Further, it has served as the fundamental basis for developing essential methods for identifying genomic tracts of adaptive change, predicting functionally important parts of the human genome, timing the origins of novel genes and mutations, and identifying genomic elements associated with human diseases (Nei and Kumar 2000; Felsenstein 2004; Yang 2006; Lynch 2007; Kumar et al. 2011).

Here, we first introduce the primary tenets of the NTME as they apply to population and species differences, and then review the evidence that these tenets explain the nature and emergence of disease-associated variants in modern human populations. We also elaborate on the clinical implications of this neutral theory perspective, which integrates population and species evolution patterns for better diagnosis and treatment of human diseases with an etiological basis in genetic variation.

⁸Corresponding author
E-mail s.kumar@asu.edu

Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.133702.111>.

Neutral theory for personal, population, and species differences

Under the NTME, the primary forces that give rise to and maintain variation within populations are mutation, recombination, and random genetic drift (for reviews, see Nei 1987; Lynch 2007; Nei et al. 2010). Mutations are de novo genomic alterations in an individual which, when passed on to offspring through the germline (i.e., sperm and eggs), may contribute to disease and appear as disease-associated variants in contemporary populations. Mutations also occur in somatic cells and mitochondrial DNA, contributing to many monogenic and complex diseases (Taylor and Turnbull 2005; Erickson 2010). Recombination breaks up associations between alleles that co-occur on the same chromosome, which can lead to or dissociate detrimental juxtapositions of allelic variants (Keightley and Otto 2006).

Kimura (1983) and Nei et al. (2010) formalized and advocated the NTME, which asserts that random genetic drift dictates the fate of a newly arisen mutation over time in a population, and that a majority of (personal, population, or species) genetic differences observed today are selectively neutral. In a population with N randomly mating diploid individuals, a novel mutation will have a population frequency of $1/(2N)$. If the mutation is selectively neutral, then it will have a probability of $1/(2N)$ to become fixed in a population, and genetic drift will govern changes in its frequency. If a mutation is selectively deleterious, then the expected population frequency in subsequent generations is less than the current population frequency due to the effects of purifying selection. Therefore, de novo mutations and variations with significant negative consequences will not rise to high frequencies, and they will often lead to diseases that affect fecundity depending on the age of their onset. Consequently, these variants will be observed in low frequencies within populations. Empirical support for these tenets is found abundantly in the observed patterns of genetic polymorphism in diverse population genomic data. For example, an analysis of dense population sequencing data reveals that the forces of purifying selection and neutral variation alone are able to explain overall genome-wide patterns of genetic variation (e.g., Hernandez et al. 2011; Lohmueller et al. 2011). Functionally important regions of the genome show much lower degrees of genetic variation than others with no known function (e.g., coding versus noncoding; Fig. 1A; Goode et al. 2010; Mu et al. 2011). Exonic variants that have become common in the population (frequency $\geq 5\%$) are overabundantly synonymous (do not change the amino acid encoded). This is because the absolute numbers of such synonymous variants exceed nonsynonymous (amino acid altering) variants even though the number of positions that can receive nonsynonymous mutations is three times greater than the positions that can receive the synonymous mutations (The 1000 Genomes Project Consortium 2010). Furthermore, rare variants (frequency $<1\%$) in exonic regions are enriched for deleterious, nonsynonymous variants (Fig. 1B; Li et al. 2010; Marth et al. 2011).

Within the NTME framework, adaptive (beneficial) mutations have a higher probability of being retained in the population, depending on the selective advantage they confer (Eyre-Walker and Keightley 2009). However, adaptive variants are expected to occur less frequently in a population in comparison with neutral polymorphisms (Kimura 1983; Boyko et al. 2008; Eyre-Walker and Keightley 2009). This occurs despite the fact that any adaptive variant has a much higher chance of rising to high frequencies. We estimate that the proportion of high-frequency variants with

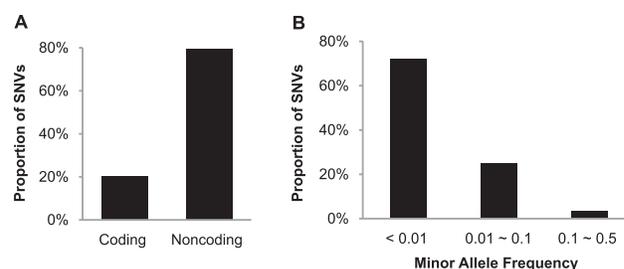


Figure 1. Observed patterns of population polymorphisms that are consistent with the neutral theory of molecular evolution. (A) Proportions of single nucleotide variants (SNVs) with a rejected substitution (RS) score greater than 3 in coding and noncoding regions (Goode et al. 2010). A lower degree of genetic variation is permitted by purifying selection in functionally important regions (coding) of the genome as compared with noncoding regions with no known function. (B) Proportions of non-synonymous SNVs predicted to be damaging by SIFT and PolyPhen-2 programs (Marth et al. 2011), which show an enrichment of deleterious (nonsynonymous) variants among the rare polymorphisms (allele frequency $<1\%$).

neutral origins in a population will far exceed, by an order of magnitude, the proportion of high-frequency adaptive variants, even under assumptions that favor adaptive forces (Fig. 2). Consequently, a vast majority of high-frequency variants observed in modern populations are more likely to be selectively neutral.

In fact, even adaptive variants with significant fitness effects may follow a fate similar to that of selectively neutral mutations if the effective population size (N_e) is small. This happens because the effectiveness of selection for a mutation is a product of fitness effect, represented by the selection coefficient (s), and N_e . Mutations with $N_e s \ll 1$ act as “effectively” neutral. Therefore, in populations with smaller N_e , the effects of random genetic drift can more easily overcome the effects of selection (Welch et al. 2008; Hurst 2009). Estimates of N_e for human populations (~ 3000 – $10,000$) are orders of magnitude less than the size of the modern human mating population (Tenesa et al. 2007), which is likely due to the occurrence of one or more population bottleneck events during early human migration (Voight et al. 2005; Lohmueller et al. 2008). This means that beneficial mutations need to have a rather high selective advantage in order to gain a significantly higher probability of rising to high frequencies due to their adaptive effects. Such facts underlie the original statement that “the essential part of the neutral theory is not so much that molecular mutants are selectively neutral in the strict sense as that their fate is largely determined by random drift” (Kimura 1983, p. 34). This also means that the original formulation of the neutral theory not only deals with the presence of deleterious mutations that are subject to negative selection, but also embraces adaptive and nearly neutral mutations (Ohta 1992; Hurst 2009), many of whose fate will be largely determined by random genetic drift (Nei et al. 2010).

From population variation to species differences

Under the NTME, population polymorphisms represent the transient phase of evolution among species (Kimura 1983). The probability that a strictly neutral variant will become fixed in a population, and thus appear as a difference between species, is equal to the rate of mutation (Kimura 1968). The functional importance and characteristics of a locus dictates the strength of purifying selection against mutations occurring at that position. This means that a position will be more conserved between species

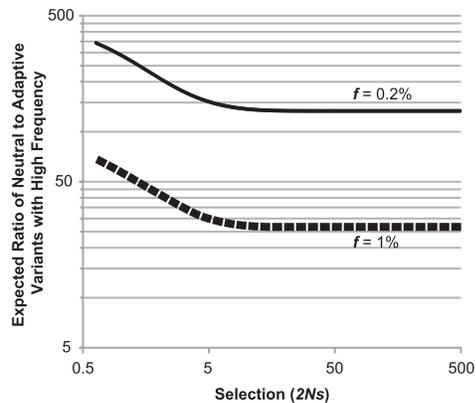


Figure 2. Relative occurrence of neutral and adaptive mutations that will reach high frequency (HF: 50%–90%) in a population, assuming that a fraction f of all new mutations is adaptive. An overwhelming majority of HF mutations in a population are expected to be neutral even when the adaptive mutations have a very high selective advantage (s , increment of fitness) relative to the wild-type allele. The fraction of HF adaptive mutations is not expected to exceed 3.5% even when 1% of all new mutations in a population are adaptive, and the adaptive allele has a very high selective advantage. The expected proportion of HF neutral and adaptive polymorphisms was estimated by using Sawyer and Hartl (1992) equations 11 and 12, which describe the equilibrium flux of mutant alleles reaching high frequency in an effective population size of $2N$. Note that increasing the $2Ns$ will not increase the proportion of adaptive HF polymorphisms, as the elevated probability of entering the high-frequency range is balanced by shortening of the transient polymorphic period. Also, we have used a rather liberal upper bound on the relative proportion of adaptive mutations ($f = 1\%$; dashed line), but the actual fraction is expected to be smaller ($f = 0.2\%$; solid line). For simplicity, only neutral and adaptive mutations were considered in the above calculations, because deleterious mutations are unlikely to rise to very high frequencies (>50%) on their own or by hitchhiking with adaptive variation, because only a small portion of the genome is hitchhiking by adaptive substitutions (Chun and Fay 2011). In any case, the maximum number of deleterious variants with high frequency will be similar to that of beneficial variants. Also, some low-frequency neutral variants will hitchhike to high frequency, which will cause a temporary increase of a high-frequency neutral derived allele during or immediately after the fixation of the adaptive mutation. This will not have any impact on the overall expectations, because such increases are followed by a long period of below-average abundance of high-frequency variants (Kim and Stephan 2000).

if substitutions between species occur slower than what would be expected under strict neutrality. This fact has been exploited in functional genomics extensively, where deviations from neutral evolutionary expectations serve as a major tool for identifying genomic regions of functional importance. If the interspecific divergence at a position is less than would be expected under neutrality, it suggests that the site is likely functionally important, and that purifying selection has prevented alternative alleles which disrupt function from reaching fixation (Goode et al. 2010).

The degree of purifying selection reflected in the interspecies evolutionary rate of substitution at a position shows a strong relationship with the population frequencies of alleles at polymorphic loci (Fig. 3; Subramanian and Kumar 2006; Kumar et al. 2009). Similarly, 10-fold more synonymous substitutions (per base pair) have occurred between human and chimpanzee proteins than nonsynonymous substitutions, a trend that is universally observed among closely and distantly related species (Clark et al. 2003; Subramanian and Kumar 2006). Furthermore, the magnitude of biochemical difference in amino acid variation within a population is very similar to that observed among amino acids

substituted between species (Subramanian and Kumar 2006). And the degree of sequence divergence at fourfold-degenerate sites is similar to that of intronic regions and pseudo-genes, because their mutations are not generally expected to have any significant impact on function (Zheng et al. 2007).

In the above, we have outlined neutral theory as it applies to the medically relevant genomic variation discussion below. A more detailed review of the biological and mathematical aspects of the neutral theory in the genomics era is found in Nei et al. (2010).

Neutral theory's implications for medically relevant variants

All population, personal, and disease-associated variants observed today have been subjected to, and their population frequencies a product of, the evolutionary processes of genetic drift and natural selection. Their unification under the NTME provides the theoretical basis upon which hypotheses about the distribution of these mutations—including simple, complex, mitochondrial, and somatic disease-associated mutations—can be developed and evaluated.

Evolutionary distributions of disease-associated variants

The NTME predicts that putatively causal disease variants in functional regions (e.g., protein-coding regions) would be observed disproportionately at slow evolving positions in the genome, which are constrained by purifying selection against harmful mutations. Empirical studies using large samples of Mendelian disease-associated and neutral variation data support these expectations (Miller and Kumar 2001; Blekhman et al. 2008; Kumar et al. 2011). A vast majority of nonsynonymous single nucleotide variants (nSNVs) associated with Mendelian diseases occur at highly evolutionarily constrained positions, whereas nSNVs found in healthy individuals occur in more variable positions (Fig. 4A).

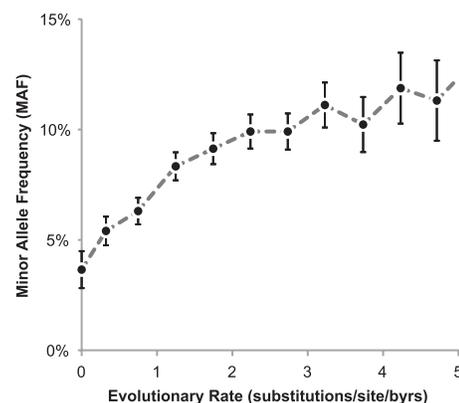


Figure 3. Relationship between the cross-species evolutionary rate and minor allele frequency (MAF) for nSNVs. Slower-evolving positions host nSNVs with the lower MAFs, on average. MAFs for 15,043 neutral nSNVs from the HumVar data set were obtained for one HapMap population (CEU) (The International HapMap 3 Consortium 2010). Long-term evolutionary rates were estimated using the protein sequence alignment from the UCSC Genome Browser resource (Kumar et al. 2009; Fujita et al. 2011). For each evolutionary rate category, the median MAF is plotted along with error bars (2 SEM).

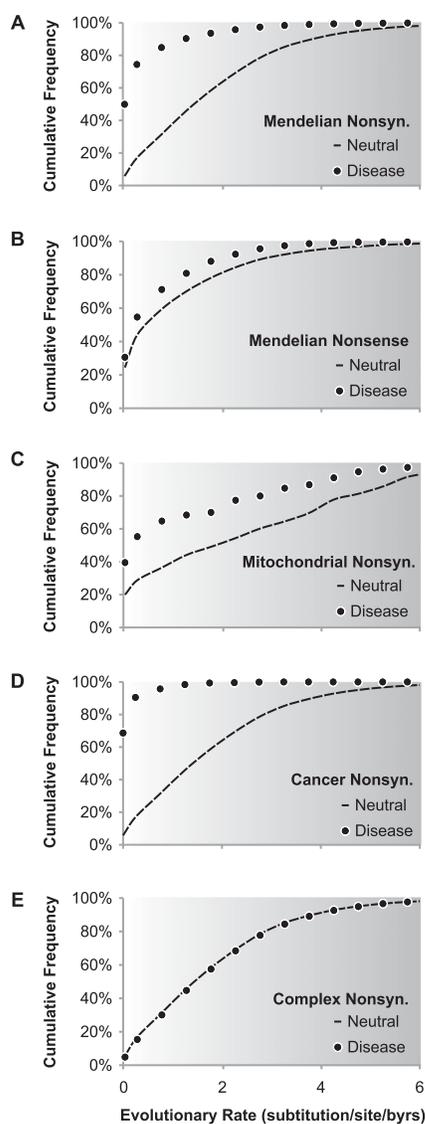


Figure 4. The cumulative frequency distribution of disease-associated variants (closed circles) and population polymorphisms (dashed lines) over long-term evolutionary substitution rates, with darker background indicating increasing evolutionary variability. (A) Nonsynonymous variants associated with Mendelian diseases; 39,578 variants from HGMD are shown (Stenson et al. 2009). (B) Nonsense variants associated with Mendelian diseases; 2806 variants from HGMD are shown. (C) Nonsynonymous variants in human mitochondrion associated with diseases; 190 disease-associated and 964 polymorphic variants from MITOMAP are shown (Ruiz-Pesini et al. 2007). (D) Nonsynonymous somatic mutations associated with cancers; 1207 driver variants from CanPredict are shown (Kaminker et al. 2007). (E) Nonsynonymous variants associated with complex diseases; 8644 variants from VARIMED are shown (Chen et al. 2010). These distributions show that protein positions harboring disease-associated variants tend to evolve at a much slower rate (lighter background) than positions harboring population polymorphisms. This pattern is the most distinct for nonsynonymous variants associated with Mendelian diseases, cancers, or mitochondrial diseases. The distinction becomes less obvious for nonsense variants, and the pattern completely disappears for nonsynonymous variants associated with complex diseases, the latter of which are expected to exhibit a neutral evolutionary pattern due to their modest effects on fecundity. Evolutionary rates in panels A, B, D, and E are estimated using multiple alignments of orthologs from 46 species (Fujita et al. 2011) following the procedure in Kumar et al. (2009). For panel C, amino acid sequence alignments for mitochondrial proteins were obtained for 28 mammalian species from MamMiBase (Vasconcelos et al. 2005) and the evolutionary rate was estimated following Kumar et al. (2009).

Missense variants that introduce premature stop codons (i.e., nonsense variants) also occur with a greater preponderance at highly conserved positions when they are associated with diseases (Fig. 4B; see also Zia and Moses 2011). However, the attenuating effects of various cellular processes (e.g., nonsense mediated decay) may enable the persistence of nonsense mutations at relatively conserved positions (Frischmeyer and Dietz 1999; Holbrook et al. 2004). Disease mutations caused by defects in the mitochondrial proteins also show a similar pattern (Fig. 4C; see also Bhardwaj et al. 2009; Montoya et al. 2009). Somatic cell mutations implicated as having a causal role in the pathogenesis of cancer (i.e., driver mutations) exhibit a trend similar to Mendelian disease variants (Fig. 4D; Kaminker et al. 2007; Forbes et al. 2008). This corresponds with the observation that driver mutations are preponderant at protein functional sites (Dixit et al. 2009; Izarzugaza et al. 2009; Talavera et al. 2010), and that they occur in regions that have significant consequences for germline fitness in addition to the somatic fitness of clonally expanding neoplastic cell populations (Fischer et al. 2011). This is different from noncancerous complex diseases, where the genetic basis of each disease is attributable to many causal variants with very small fitness effects. In accordance with the expectations of NTME, nSNVs implicated in complex diseases with modest fitness effects exhibit evolutionary trends similar to neutral nSNVs (Fig. 4E; see also Thomas and Kejariwal 2004).

Since evolutionary patterns are shaped by reproductive success, we would expect late-onset diseases, which manifest after reproductive age, to leave a weaker evolutionary imprint. Therefore, we would expect smaller differences between neutral variants and variants involved in late-onset diseases. This is indeed the case, as the amino acid variants involved in late-onset diseases show greater biochemical severities than those implicated in early-onset Mendelian diseases (Vitkup et al. 2003; Subramanian and Kumar 2006). Additionally, mutations implicated in complex diseases, many of which manifest their most serious debilitating effects after reproductive age, occur at evolutionarily less conserved positions as compared with Mendelian disease mutations (Wright et al. 2003; Thomas and Kejariwal 2004; Kumar et al. 2011).

Variants observed in personal exomes also follow the expectations of the NTME. Novel variants arising by de novo mutations in functionally important and, thus, slowly evolving positions in an exome are subjected to a greater degree of (purifying) selection. Consequently, the numbers of variants observed at evolutionarily constrained positions in a personal exome are much smaller than expected (observed/expected ratio <1), with an increasingly higher overabundance of nSNVs at faster evolving positions (Fig. 5). These footprints of neutral evolutionary processes in personal exomes ultimately shape the allele frequencies of nSNVs in populations (Fig. 3). Furthermore, it is expected that the balance between mutations and purifying selection across the human genome will result in rare minor alleles at the majority of polymorphic sites, which is supported by the analyses of whole genome and exome sequence data measured from diverse human populations (Gravel et al. 2011; Lohmueller et al. 2011). This would mean that the majority of disease-associated variations in the human population are likely due to low frequency and rare alleles, and these variants are expected to contribute disproportionately to functional population variance under the NTME (Pritchard 2001; Pritchard and Cox 2002; Marth et al. 2011; Zhu et al. 2011; MacArthur et al. 2012). Recent studies based on evolutionary modeling and population sequencing confirm this expectation (Dickson et al. 2010; Eyre-Walker 2010; Gravel et al. 2011; Marth et al. 2011; Zhu et al.

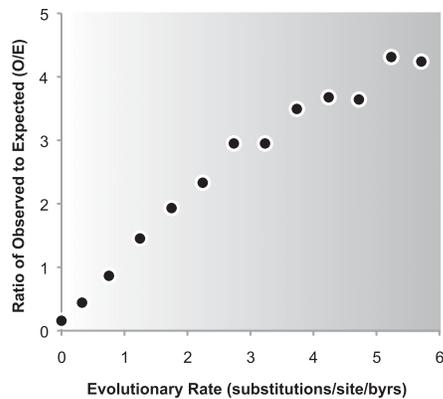


Figure 5. The ratio of observed to expected (O/E) numbers of personal exome variants (nonsynonymous, nSNVs) found in positions with higher to lower degrees of evolutionary conservation (slow to fast evolutionary rates). There is an underabundance of nSNVs in slow evolving positions ($O/E < 1$; lighter background) and an overabundance at fast evolving positions ($O/E > 1$; darker background). A linear regression model fits the observed data well ($R^2 > 0.95$), which denotes a decrease in the strength of natural selection with the increasing evolutionary rate, as expected under the neutral theory. Sequences of eight HapMap exomes (Ng et al. 2009) were used to generate the figure, based on evolutionary rates estimated using 46 species alignment of vertebrates and lamprey from the UCSC Genome Browser resource (Kumar et al. 2009; Fujita et al. 2011).

2011) and may explain some of the inability of current genetic association studies to detect many causal variants, as they are typically not sufficiently powered to detect associations with rare alleles at a majority of the polymorphic sites (Marth et al. 2011).

Adaptive histories of disease variants

Many specific cases of the interplay between evolution and human health have been focused on a few detrimental mutations that are found in appreciably higher frequencies in modern human populations (Table 1). Such high-frequency disease-associated variants can result from past adaptive events, which have become detrimental in the modern contexts. For example, a long-held adaptation-oriented view on the evolutionary origins of obesity draws from the “thrifty gene” hypothesis, which suggests that metabolic adaptation was driven by periods of food scarcity in hunter-gatherer populations (Neel 1962). Such views, and a few individual disease-associated variants with seemingly compelling adaptive explanations (Table 1), have elevated adaptation as a popular evolutionary modality for the genome-wide discovery of novel trait-associated loci in the human genome (Hindorff et al. 2009; López Herráez et al. 2009; Grossman et al. 2010; Wu and Zhang 2010).

However, such attempts to link many human diseases to adaptive evolutionary events face numerous problems. First, many signatures of genomic adaptation in humans are expected to be attributable to specific local environments, e.g., cattle domestication and malaria epidemics (Tishkoff et al. 2006; Coop et al. 2009; Novembre and Di Rienzo 2009). In this case, local adaptations will not necessarily have significant implications for disease, as all subpopulations with different environments carry their own landscape of neutral alleles (Lupski et al. 2011), and the pathogenicity of local adaptations may not be realized until there is change in the local environment, or individuals become adversely “misplaced” through recent migration, e.g., sickle-cell anemia in the African-American population (Campbell and Tishkoff 2008).

Also, when an adaptive variant is in the process of spreading in a given population, nonexpression of the adaptive trait is not necessarily deleterious, because the starting population with the former wild-type allele may not have been seriously unfit. Indeed, the thrifty hypothesis has been challenged in recent population genomic and epidemiological studies (Kimura et al. 2007; Prentice et al. 2008; Speakman 2008). However, it is plausible that extreme pressures on fitness, such as emergent infectious agents, could impose strong negative selection on wild-type alleles, and may expose deleterious cryptic variation through decanalization (Gibson 2009) or hitchhiking (Chun and Fay 2011).

Selection against previously adaptive variation is only one possible evolutionary explanation for the existence of high-frequency variants associated with clinical phenotypes. Another explanation involves neutral mutations, many of which are expected to rise to high frequency within a population due to random genetic drift. Because the negative selection (and thus neutrality) of a variant depends on its genomic and environmental contexts, which can change over time and geography, many high-frequency variants with neutral origins (previously neutral) may also contribute to human diseases today. Population migration is well known to affect the neutrality of alleles through altering the landscape of population variation in subsequent migrated generations. Repeated patterns of migration and settlement can lead to reduced genetic diversity in migrated populations due to serial founder effects (DeGiorgio et al. 2011). Migration during range expansion can also affect frequency distributions, with both neutral and deleterious variants having higher probabilities of fixation on wave fronts of range expansion due to surfing effects (Klopfstein et al. 2006; Travis et al. 2007). Migration can lead to drastic changes in the environmental context, which can expose standing neutral variation to negative selection (Hancock et al. 2010). Overall, population migration and genetic admixing disrupt the neutrality of an existing variation, subjecting it to negative selection (Stajich and Hahn 2005; Nielsen et al. 2009; Wall et al. 2009). Therefore, any high-frequency variant observed to be associated with disease in modern contexts would be either previously adaptive or previously neutral in origin.

Using the population genetics aspects of the NTME, one can assess the relative contributions of previously neutral versus previously adaptive mutations and establish a priori expectations of the degree to which the (positive) natural selection is contributing to disease today (Fullerton et al. 2000). Such a calculation suggests that the proportion of high-frequency variants with neutral origins will far exceed the proportion of high-frequency variants due to adaptation, as shown earlier (Fig. 2). This observation is corroborated by recent population genomic studies, which have reported a relatively minor role for adaptation in the emergence of disease-associated variants (Boyko et al. 2008; Myles et al. 2008; Coop et al. 2009; Hofer et al. 2009). An analysis of resequenced data for more than a hundred human genomes has shown that genomic diversity around human lineage specific amino acid substitutions was no more pronounced than that around synonymous substitutions (Hernandez et al. 2011), the latter of which are not expected to be subjected to positive selection to any significant degree. Also, the genomic diversity around functional regions does not appear to be substantially different than the genomic background, nor are these regions enriched for alleles that segregate among populations (Hernandez et al. 2011). These observations suggest that classic selective sweeps for adaptive alleles may not be frequent in human evolution. At the same time, we note that the existing methods may be underpowered to detect the presence

Table 1. Some examples of genes harboring human disease mutations for which empirical molecular evidence supports a putative role for positive natural selection acting on emerged gene variants

Gene	Product	Biological process	Associated disease(s)	Reference(s)
<i>ADH</i>	Alcohol dehydrogenase	Alcohol metabolism	Alcohol dependence	Osier et al. 2002
<i>AGT</i>	Angiotensinogen	Salt regulation pathway	Hypertension	Nakajima et al. 2004; Thompson et al. 2004
<i>ALDH2</i>	Aldehyde dehydrogenase 2	Alcohol metabolism	Alcoholism	Oota et al. 2004
<i>APOE</i>	Apolipoprotein E	Lipid metabolism	Alzheimer's disease, familial Dysbetalipoproteinemia, sea-blue histiocytosis	Hanlon and Rubinsztein 1995; Gearing et al. 1996
<i>CAPN10</i>	Calpain 10	Cellular response to insulin stimulus	Type 2 diabetes	Fullerton et al. 2000; Vander Molen et al. 2005
<i>CYP3A5</i>	Cytochrome P450, family 3, subfamily A, polypeptide	Metabolism of exogenous substances	Increased blood pressure	Thompson et al. 2004
<i>DARC</i>	Duffy antigen, chemokine receptor	Receptor signaling	Malaria	Hamblin and Di Rienzo 2000
<i>DMD</i>	Dystrophin	Regulation of muscle fiber activity	Lassa fever, muscular dystrophy	Kunz et al. 2005; Sabeti et al. 2007
<i>DRD4</i>	Dopamine receptor D4	Dopamine receptor	ADHD	Ding et al. 2002; Wang et al. 2004
<i>EDAR</i>	Ectodysplasin A receptor	Hair follicle patterning	Hypohidrotic ectodermal dysplasia	Sabeti et al. 2007
<i>EPAS1/HIF2A</i>	Endothelial PAS domain protein 1	Response to hypoxia	Erythrocytosis familial type 4	Beall et al. 2010
<i>G6PD</i>	Glucose-6-phosphate dehydrogenase	Cytosolic enzyme	Protection from <i>Plasmodium falciparum</i> malaria vs. G6PD deficiency	Tishkoff et al. 2001; Saunders et al. 2002
<i>HBA, HBB, and HBS</i>	Alpha- and beta-globin	Synthesis of hemoglobin	Malaria, sickle-cell anemia	Kwiatkowski 2005
<i>HFE</i>	Hemochromatosis	Regulate iron absorption	Decreased risk of anemia in pregnancy vs. hemochromatosis	Ajioka et al. 1997; Toomajian and Kreitman 2002; Toomajian et al. 2003
<i>LARGE</i>	Mucin-like glycosyltransferase	Regulation of muscle fiber activity	Lassa fever, muscular dystrophy	Sabeti et al. 2007
<i>LCT</i>	Lactase	Lactose metabolism	Lactose intolerance	Tishkoff et al. 2006
<i>MAOA</i>	Monoamine oxidase A	Neurotransmitter	Mental retardation, Behavioral disorders	Gilad et al. 2002
<i>MCPH1</i>	Microcephalin 1	Putative roles in neurogenesis and regulation of brain size	Primary microencephaly	Evans et al. 2005; Mekel-Bobrov et al. 2005
<i>MMP3</i>	Matrix metalloproteinase 3	Breakdown of extracellular matrix	Coronary artery disease	Rockman et al. 2004
<i>OPN1LW</i>	Opsin 1 (cone pigments), long-wave-sensitive	Photoreceptor activity	Color blindness	Verrelli and Tishkoff 2004
<i>PDYN</i>	Opioid neuropeptide precursor prodynorphin	Cognition and pain response	Schizophrenia, cocaine addictions, and epilepsy	Rockman et al. 2005
<i>MEFV</i> (pyrin)	Mediterranean fever	Modulator of innate immunity	Familial Mediterranean fever	Schaner et al. 2001

of the limited degree of adaptive variation that is likely to exist, or different models of adaptive evolution in human history may remain to be discovered.

Historical trends of the human lifespan suggest that selection against previously neutral variants is likely to be frequent in modern human populations. Natural selection acts in accordance with reproductive success, which is linked to life spans of individuals in populations. Many contemporary population alleles were once neutral and reached high frequencies when fitness was determined at a shorter average life span, e.g., ~30 yr at birth (Caspari and Lee 2004). However, the average life spans of humans have grown significantly and steadily over the past centuries (Wilmoth 2000), which exposes nonneutrality of many variants only in later life stages (Wright et al. 2003). Thus, one would predict an increase in the numbers of mildly deleterious (high and low frequency) variants segregating in the population, which is supported by the recent analysis of standing genetic variation observed in whole-genome sequence data (Li et al. 2010; Zhu et al.

2011). Because previously neutral variants vastly outnumber adaptive variations (Fig. 2), we expect to discover a growing number of previously neutral variants involved in diseases (Pritchard 2001; Kryukov et al. 2007). Directional selection acting on previously neutral variants is difficult to discern from neutral variation using common statistical methods for detecting selective sweeps (Teshima et al. 2006). Nonetheless, signs of selection on previously neutral variants are already being reported, e.g., Crohn's disease risk alleles in the *NOD2* gene (Nakagome et al. 2012). Similar patterns of selection have been inferred for various traits associated with environmental adaptations (Hancock et al. 2010). Indeed, adaptive variation will underlie a small proportion of human genetic disease variation, many of which will be due to pressures from infectious disease or fitness trade-offs due to antagonistic pleiotropy (Finch 2010). However, these adaptive explanations must take the form of alternative hypotheses that should be rigorously evaluated against a neutral null hypothesis.

Neutrality has already served as a null hypothesis in the development and application of advanced approaches for detecting signals of recent positive selection (Fu and Li 1993; Sabeti et al. 2002, 2006; Voight et al. 2006; Kimura et al. 2007; Sabeti et al. 2007; Williamson et al. 2007). Notable findings have emerged from the use of such approaches, including recent selection in the lactase (*LCT*) gene that allows some human populations to digest lactose into adulthood (Tishkoff et al. 2006), and adaptations associated with response to infectious disease (Ohashi et al. 2004; Corona et al. 2010). Furthermore, Chun and Fay (2011) reported that common disease-associated variants were more abundant than rare disease-associated variants in hitchhiking regions, although both types of variants are enriched in hitchhiking regions compared with nonhitchhiking regions. As we discover more adaptive substitutions leading to the evolutionary lineage of modern humans, as well as population-specific molecular adaptations, we expect that the adaptive origins of many more variants implicated in human diseases will be revealed (Pickrell et al. 2009; Corona et al. 2010; Grossman et al. 2010; Casto and Feldman 2011). Of course, the adaptive variants may have a higher probability of becoming pathogenic in altered environments and genetic contexts as compared with previously neutral alleles. Nevertheless, we expect their absolute numbers will be overshadowed by the much greater preponderance of alleles having neutral origins.

Neutral evolution as a tool to identify medically relevant variants

Distinguishing harmful variations in personal genomes from those with benign effects during an individual's life course is a grand challenge in genomic medicine and personal genomics. Due to advances in sequencing technologies, medical practitioners are already being asked to diagnose mutations in a patient's genome in addition to their physiological symptoms. Presently, the complete genomes and exomes of thousands of individual humans have been sequenced, with projects underway to expand these numbers to tens to hundreds of thousands (Levy et al. 2007; Wang et al. 2008; Wheeler et al. 2008; Pushkarev et al. 2009; The 1000 Genomes Project Consortium 2010). Each personal genome carries thousands of protein variants and over a million noncoding DNA variants (Ng et al. 2009; Pushkarev et al. 2009; Cirulli and Goldstein 2010; Dewey et al. 2011; Kumar et al. 2011). The clinical assessment of a patient incorporating a personal genome has the promise of improving the overall appraisal of a patient's health risks and guiding clinical decision-making regarding lifestyle and therapeutic intervention (e.g., Ashley et al. 2010). Although high-throughput, genome-wide molecular profiling technologies such as RNA sequencing (RNA-seq) and regulatory profiling through chromatin immunoprecipitation (ChIP-seq) and related techniques provide powerful modalities for profiling the functional effects of individual variants (Kasowski et al. 2010), they are not yet amenable to regular clinical use, and still typically incorporate evolutionary profiling for data interpretation (MacArthur et al. 2012; Ward and Kellis 2012). In fact, experimental or clinical evidence regarding the consequence of mutations even commonly observed in the population is often lacking, and recent, high-throughput efforts to identify common disease variants through genome-wide association studies (GWAS) have so far offered limited clinical utility (Dupuis and O'Donnell 2007; McCarthy and Hirschhorn 2008; Manolio et al. 2009).

Without any means to establish prior expectations of clinical significance on rare and common variants, their incidental discovery may cause the patient to be subjected to unnecessary clinical scrutiny (Kohane et al. 2006). Thus, there is an impending clinical need to be able to sort through variants in individual human genomes, and highlight those that are most likely to have a consequence for an individual's health or course of clinical care. Fortunately, the expectations and evolutionary interpretations of mutational patterns asserted under the NTME are coherent across taxonomic levels (species, populations, and individuals). This provides a strong theoretical and empirical basis for establishing baseline expectations for the clinical assessment of population variation and personal mutations surveyed from contemporary personal genomes. Neutral expectations and the interpretation of evolutionary patterns across species can improve our ability to detect and diagnose the full spectrum of variants present in a personal genome (Kumar et al. 2011). For example, single-nucleotide estimates of cross-species evolutionary constraint have been used to identify causal disease variants for Freeman-Sheldon syndrome and Miller syndrome from exome sequencing data (Cooper et al. 2010), and similar estimates of cross-species evolutionary rate were used to filter for functional rare variants in the clinical assessment of whole genomes for a family quartet (Dewey et al. 2011).

More broadly, computational approaches have been developed over the last decade to predict whether a variant impacts an individual, especially for protein variants. These tools incorporate evolutionary conservation along with an increasingly large number of clinical or biological attributes into their algorithms (Karchin 2009; Cline and Karchin 2011). Although designed primarily for the analysis of coding mutations in candidate genes for the purpose of hypothesis generation, these tools are now being utilized for genome-wide scans of coding polymorphisms (Doniger et al. 2008; Lohmueller et al. 2008). These tools appear to perform quite well for the purpose of generating biological hypotheses (Adzhubei et al. 2010; Gonzalez-Perez and Lopez-Bigas 2011). However, they frequently produce discordant diagnoses even when scanning for simple (monogenic) diseases (e.g., Chun and Fay 2009), indicating a need to improve their accuracy for the purpose of clinical diagnosis. While most computational tools already incorporate evolutionary information in their algorithms (Ng and Henikoff 2006), long-term evolutionary histories of species (that largely capture neutral evolutionary trends over millennia) can provide information to forecast when the computational predictions are expected to be correct (Kumar et al. 2009, 2011). For example, the computational tools are found to be least accurate to correctly diagnose disease variation at the fastest-evolving positions (Kumar et al. 2009), which would be predicted to harbor mostly benign mutations under the NTME (Kimura 1983; Subramanian and Kumar 2006).

Another emerging application of evolutionary information is their use as priors in population-based studies of human complex diseases having an etiological basis in human mutation. Evolutionary profiles of causal complex disease mutations are very similar to those of neutral population polymorphisms (Fig. 4E; Thomas and Kejariwal 2004; Blekhman et al. 2008; Cai et al. 2009), because complex disease variants typically have modest effects on fecundity, and also often have substantial nongenetic components through environmental influences. Therefore, it is frequently thought that evolutionary estimates of constraint and neutrality are not very informative for identifying complex disease variation.

Consequently, modern GWAS of complex disease traits operate under the statistical assumption that all measured loci are equally likely to harbor alleles that segregate between healthy and disease-affected populations. However, we know from the empirical data and established evolutionary principals that the expected population allele frequency spectra for both ancient and derived alleles are tied to the evolutionary history of positions, and that certain categories of disease mutations have a propensity to be found at sites with particular evolutionary histories (Cooper et al. 2010; Kumar et al. 2011).

Indeed, it is now clear that long-term evolutionary priors estimated from multispecies alignments can improve the chances of discovering reproducible variants in genetic disease association studies (Dudley et al. 2012). Because of the large difference in the depths of the human population coalescence tree (~100,000 yr) and the time to the most recent common ancestor with another species (5–7 million years; Kumar and Hedges 2011), long-term (i.e., across species) evolutionary quantities can be generally estimated for each position in the human genome independent of human populations. Their use in identifying reproducible variants that explain greater proportions of genetic heritability of common diseases helps to address the issue of “missing” heritability of complex diseases (Manolio et al. 2009; Dudley et al. 2012). Therefore, it is likely that new types of association statistics that incorporate evolutionary priors would have increased the power to discover reliable disease-associated loci from both published and future genetic disease association studies (Dudley et al. 2012).

At this point, it is important to highlight some other important limitations in the use of evolutionary information as a tool for diagnosing mutations. For example, as mentioned earlier, a historically neutral allele may be rendered nonneutral in a modern population context due to changes in environmental exposures (previously neutral). Additionally, mutations that may have no selective impact on fecundity may exhibit real deleterious effects subsequent to reproductive age, thereby escaping the filter of purifying selection across species. In such cases, long-term evolutionary trends are unlikely to inform on the disease propensity of variants. It is also possible that compensatory mutations could render previously deleterious variants neutral, causing them to revert to their deleterious nature under the context of a different genetic background (Kondrashov et al. 2002; Kern and Kondrashov 2004). This complicates neutral interpretations, as compensatory mutation events can be difficult to detect using existing methods. However, compensatory mutations would confound interpretation of any type of context-dependent selection; thus, neutral and adaptive interpretations would be equally affected.

Nevertheless, the useful application of molecular evolution to medicine has advanced beyond the contemporary discussions in evolutionary (and Darwinian) medicine of health and disease that often place a primary focus on human population histories and historical environments in understanding root causes (Nesse and Williams 1996; Gluckman and Hanson 2004; Stearns et al. 2010). Indeed, the ultimate achievement of a null hypothesis for human disease mutations founded in the NTME is to unify evolutionary biology, genomics, and modern evidence-based medicine. It is clear from the above discussion that evolution serves not as a pathological ghost of human history, but rather as a powerful framework for translating fundamental discoveries into clinical utility toward improved patient outcomes.

A healthy perspective on evolution in molecular medicine

It is important to address the role that popular perceptions might play in the acceptance of evolutionary thinking in medicine (Nesse and Williams 1996; Ewald 2000). Perhaps the most unfortunate consequence of a purely adaptation-oriented view of human disease mutations is the belief that evolution serves as a wellspring for many facets of ill health. Some protagonists of this view have even been so bold as to declare which specific “failures” of evolutionary adaptation have caused specific diseases in modern populations (Boaz 2002). Such popular views appear to be inspired by the early adaptive explanations of human diseases (Allison 1954), where a common-variant-common-disease model was proposed to explain genetic diseases (e.g., Pritchard and Cox 2002), based on an assumption that high-frequency variants arise purely due to adaptive advantages. Of course, changes in environment will expose some adaptive variants carried by many individuals to negative selection and, thus, disease. However, environmental changes will also disrupt the neutrality of many variants that have reached high frequency by random genetic drift, without adaptation. In a population, high-frequency neutral variants vastly outnumber adaptive variants, and the contribution of these neutral alleles to the common-variant-common-disease scenario has not received attention historically. In fact, an NTME based explanation does not require one to choose between rare-variant and common-variant hypotheses for common diseases, as their relative contributions will be a function of the timing of the environmental changes and the allele frequency landscapes of the contributing genes and genomic segments. Furthermore, selection will act on individual rare and common variants in the broader context of their joint functional effects across biological pathways (Gibson 2012).

Overall, a neutral view suggests that historical adaptive evolutionary events are not the source of illness, but rather that evolution is the source of robustness and the reason that humans thrive so successfully under a broad range of conditions. It rejects the minority assertion that evolution has somehow “stopped” for humans because modern medicine and lifestyles have slowed adaptation (Belluz 2008; Stearns et al. 2010). A neutral perspective asserts, with overwhelming support from existing data, that the drumbeat of evolution continues in modern populations irrespective of differences in the fitness landscapes between modern and ancient human populations. In this way, evolution gains relevance on par with other continuous factors influencing individual health, such as behavior and environment. Finally, a crucial step toward integrating evolution and medicine rests in orienting empirical findings from evolutionary studies such that they are easily translated toward clinical application.

Acknowledgments

We thank Vanessa Gray and anonymous reviewers for helpful comments, and Carol Williams for support in editing the manuscript. J.T.D. was supported by the NLM Biomedical Informatics Training Grant to Stanford University (T15 LM007033). Y.K. is supported by Ewha Global Top 5 Grant 2011 of Ewha Womans University. A.J.B. is supported by funding from NIH/NIGMS (R01 LM009719). S.K. is supported by funding from NIH/NLM (R01 LM010834) and NIH/NHGRI (R01 HG002096).

References

- The 1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467**: 1061–1073.
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* **7**: 248–249.
- Ajioka RS, Jorde LB, Gruen JR, Yu P, Dimitrova D, Barrow J, Radisky E, Edwards CQ, Griffen LM, Kushner JP. 1997. Haplotype analysis of hemochromatosis: Evaluation of different linkage-disequilibrium approaches and evolution of disease chromosomes. *Am J Hum Genet* **60**: 1439–1447.
- Allison AC. 1954. Protection afforded by sickle-cell trait against subtertian malarial infection. *Br Med J* **1**: 290–294.
- Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, et al. 2010. Clinical assessment incorporating a personal genome. *Lancet* **375**: 1525–1535.
- Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, Li C, Li JC, Liang Y, McCormack M, et al. 2010. Natural selection on *EPAS1* (*HIF2 α*) associated with low hemoglobin concentration in Tibetan highlanders. *Proc Natl Acad Sci* **107**: 11459–11464.
- Belluz J. 2008. Leading geneticist Steve Jones says human evolution is over. *The Times* (London).
- Bhardwaj A, Mukerji M, Sharma S, Paul J, Gokhale CS, Srivastava AK, Tiwari S. 2009. MtSNPscore: A combined evidence approach for assessing cumulative impact of mitochondrial variations in disease. *BMC Bioinformatics* (Suppl 8) **10**: S7. doi: 10.1186/1471-2105-10-S8-S7.
- Blekhman R, Man O, Herrmann L, Boyko AR, Indap A, Kosiol C, Bustamante CD, Teshima KM, Przeworski M. 2008. Natural selection on genes that underlie human disease susceptibility. *Curr Biol* **18**: 883–889.
- Boaz NT. 2002. *Evolving health: The origins of illness and how the modern world is making us sick*. Wiley, New York.
- Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, et al. 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* **4**: e1000083. doi: 10.1371/journal.pgen.1000083.
- Cai JJ, Borenstein E, Chen R, Petrov DA. 2009. Similarly strong purifying selection acts on human disease genes of all evolutionary ages. *Genome Biol Evol* **1**: 131–144.
- Campbell MC, Tishkoff SA. 2008. African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* **9**: 403–433.
- Caspari R, Lee SH. 2004. Older age becomes common late in human evolution. *Proc Natl Acad Sci* **101**: 10895–10900.
- Casto AM, Feldman MW. 2011. Genome-wide association study SNPs in the human genome diversity project populations: Does selection affect unlinked SNPs with shared trait associations? *PLoS Genet* **7**: e1001266. doi: 10.1371/journal.pgen.1001266.
- Chen R, Davydov EV, Sirota M, Butte AJ. 2010. Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS ONE* **5**: e13574. doi: 10.1371/journal.pone.0013574.
- Chun S, Fay JC. 2009. Identification of deleterious mutations within three human genomes. *Genome Res* **19**: 1553–1561.
- Chun S, Fay JC. 2011. Evidence for hitchhiking of deleterious mutations within the human genome. *PLoS Genet* **7**: e1002240. doi: 10.1371/journal.pgen.1002240.
- Cirulli ET, Goldstein DB. 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* **11**: 415–425.
- Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, et al. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**: 1960–1963.
- Cline MS, Karchin R. 2011. Using bioinformatics to predict the functional impact of SNVs. *Bioinformatics* **27**: 441–448.
- Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, Absher D, Myers RM, Cavalli-Sforza LL, Feldman MW, Pritchard JK. 2009. The role of geography in human adaptation. *PLoS Genet* **5**: e1000500. doi: 10.1371/journal.pgen.1000500.
- Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, Shendure J, Nickerson DA. 2010. Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat Methods* **7**: 250–251.
- Corona E, Dudley JT, Butte AJ. 2010. Extreme evolutionary disparities seen in positive selection across seven complex diseases. *PLoS ONE* **5**: e12236. doi: 10.1371/journal.pone.0012236.
- DeGiorgio M, Degnan JH, Rosenberg NA. 2011. Coalescence-time distributions in a serial founder model of human evolutionary history. *Genetics* **189**: 579–593.
- Dewey FE, Chen R, Cordero SP, Ormond KE, Caleshu C, Karczewski KJ, Whirl-Carrillo M, Wheeler MT, Dudley JT, Byrnes JK, et al. 2011. Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. *PLoS Genet* **7**: e1002280. doi: 10.1371/journal.pgen.1002280.
- Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. 2010. Rare variants create synthetic genome-wide associations. *PLoS Biol* **8**: e1000294. doi: 10.1371/journal.pbio.1000294.
- Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM, et al. 2002. Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proc Natl Acad Sci* **99**: 309–314.
- Dixit A, Yi L, Gowthaman R, Torkamani A, Schork NJ, Verkhivker GM. 2009. Sequence and structure signatures of cancer mutation hotspots in protein kinases. *PLoS ONE* **4**: e7485. doi: 10.1371/journal.pone.0007485.
- Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang SP, Fay JC. 2008. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet* **4**: e1000183. doi: 10.1371/journal.pgen.1000183.
- Dudley JT, Chen R, Sandorford M, Butte AJ, Kumar S. 2012. Evolutionary meta-analysis of association studies reveals ancient constraints affecting disease marker discovery. *Mol Biol Evol* (in press). doi: 10.1093/molbev/mss079.
- Dupuis J, O'Donnell CJ. 2007. Interpreting results of large-scale genetic association studies: Separating gold from fool's gold. *JAMA* **297**: 529–531.
- Erickson RP. 2010. Somatic gene mutation and human disease other than cancer: An update. *Mutat Res* **705**: 96–106.
- Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson JR, Vaez-Azizi LM, Tishkoff SA, Hudson RR, Lahn BT. 2005. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* **309**: 1717–1720.
- Ewald PW. 2000. *Plague time: How stealth infections cause cancers, heart disease, and other deadly ailments*. Free Press, New York.
- Eyre-Walker A. 2010. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc Natl Acad Sci* (Suppl 1) **107**: 1752–1756.
- Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol* **26**: 2097–2108.
- Feero WG, Guttmacher AE, Rotimi CN, Jorde LB. 2010. Ancestry and disease in the age of genomic medicine. *N Engl J Med* **363**: 1551–1558.
- Felsenstein J. 2004. *Inferring phylogenies*. Sinauer Associates, Sunderland, Mass.
- Finch CE. 2010. Evolution in health and medicine Sackler colloquium: Evolution of the human lifespan and diseases of aging: Roles of infection, inflammation, and nutrition. *Proc Natl Acad Sci* (Suppl 1) **107**: 1718–1724.
- Fischer A, Greenman C, Mustonen V. 2011. Germline fitness-based scoring of cancer mutations. *Genetics* **188**: 383–393.
- Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR. 2008. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* **57**: 10.11.1–10.11.26.
- Frischmeyer PA, Dietz HC. 1999. Nonsense-mediated mRNA decay in health and disease. *Hum Mol Genet* **8**: 1893–1900.
- Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, et al. 2011. The UCSC Genome Browser database: Update 2011. *Nucleic Acids Res* **39**: D876–D882.
- Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, Stengard JH, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, et al. 2000. Apolipoprotein E variation at the sequence haplotype level: Implications for the origin and maintenance of a major human polymorphism. *Am J Hum Genet* **67**: 881–900.
- Gearing M, Tigges J, Mori H, Mirra SS. 1996. A β 40 is a major form of β -amyloid in nonhuman primates. *Neurobiol Aging* **17**: 903–908.
- Gibson G. 2009. Decanalization and the origin of complex disease. *Nat Rev Genet* **10**: 134–140.
- Gibson G. 2012. Rare and common variants: Twenty arguments. *Nat Rev Genet* **13**: 135–145.
- Gilad Y, Rosenberg S, Przeworski M, Lancet D, Skorecki K. 2002. Evidence for positive selection and population structure at the human MAO-A gene. *Proc Natl Acad Sci* **99**: 862–867.
- Gluckman P, Hanson M. 2004. Living with the past: Evolution, development, and patterns of disease. *Science* **305**: 1733–1736.
- Gluckman PD, Beddle A, Hanson MA. 2009. *Principles of evolutionary medicine*. Oxford University Press, Oxford.

- Gonzalez-Perez A, Lopez-Bigas N. 2011. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* **88**: 440–449.
- Goode DL, Cooper GM, Schmutz J, Dickson M, Gonzales E, Tsai M, Karra K, Davydov E, Batzoglou S, Myers RM, et al. 2010. Evolutionary constraint facilitates interpretation of genetic variation in resequenced human genomes. *Genome Res* **20**: 301–310.
- Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, Project TG, Bustamante CD. 2011. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci* **108**: 11983–11988.
- Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O, et al. 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* **327**: 883–886.
- 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.
- Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, Gebremedhin A, Sukernik R, Utermann G, Pritchard J, Coop G, et al. 2010. Colloquium paper: Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proc Natl Acad Sci (Suppl 2)* **107**: 8924–8930.
- Hanlon CS, Rubinsztein DC. 1995. Arginine residues at codons 112 and 158 in the apolipoprotein E gene correspond to the ancestral state in humans. *Atherosclerosis* **112**: 85–90.
- Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, Mcvean G, Sella G, Przeworski M. 2011. Classic selective sweeps were rare in recent human evolution. *Science* **331**: 920–924.
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci* **106**: 9362–9367.
- Hofer T, Ray N, Wegmann D, Excoffier L. 2009. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Ann Hum Genet* **73**: 95–108.
- Holbrook JA, Neu-Yilik G, Hentze MW, Kulozik AE. 2004. Nonsense-mediated decay approaches the clinic. *Nat Genet* **36**: 801–808.
- Hurst LD. 2009. Fundamental concepts in genetics: Genetics and the understanding of selection. *Nat Rev Genet* **10**: 83–93.
- The International HapMap 3 Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**: 52–58.
- Izarzugaza JM, Redfern OC, Orengo CA, Valencia A. 2009. Cancer-associated mutations are preferentially distributed in protein kinase functional sites. *Proteins* **77**: 892–903.
- Kaminker JS, Zhang Y, Waugh A, Haverty PM, Peters B, Sebisanoovic D, Stinson J, Forrest WF, Bazan JF, Seshagiri S, et al. 2007. Distinguishing cancer-associated missense mutations from common polymorphisms. *Cancer Res* **67**: 465–473.
- Karchin R. 2009. Next generation tools for the annotation of human SNPs. *Brief Bioinform* **10**: 35–52.
- Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM, Habegger L, Rozowsky J, Shi M, Urban AE, et al. 2010. Variation in transcription factor binding among humans. *Science* **328**: 232–235.
- Keightley PD, Otto SP. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**: 89–92.
- Kern AD, Kondrashov FA. 2004. Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. *Nat Genet* **36**: 1207–1212.
- Kim Y, Stephan W. 2000. Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics* **155**: 1415–1427.
- Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* **217**: 624–626.
- Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.
- Kimura R, Fujimoto A, Tokunaga K, Ohashi J. 2007. A practical genome scan for population-specific strong selective sweeps that have reached fixation. *PLoS ONE* **2**: e286. doi: 10.1371/journal.pone.0000286.
- Klopfstein S, Currat M, Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* **23**: 482–490.
- Kohane IS, Masys DR, Altman RB. 2006. The incidentalome: A threat to genomic medicine. *JAMA* **296**: 212–215.
- Kondrashov AS, Sunyaev S, Kondrashov FA. 2002. Dobzhansky-Muller incompatibilities in protein evolution. *Proc Natl Acad Sci* **99**: 14878–14883.
- Kryukov GV, Pennacchio LA, Sunyaev SR. 2007. Most rare missense alleles are deleterious in humans: Implications for complex disease and association studies. *Am J Hum Genet* **80**: 727–739.
- Kumar S, Hedges SB. 2011. TimeTree2: Species divergence times on the iPhone. *Bioinformatics* **27**: 2023–2024.
- Kumar S, Suleski MP, Markov GJ, Lawrence S, Marco A, Filipowski AJ. 2009. Positional conservation and amino acids shape the correct diagnosis and population frequencies of benign and damaging personal amino acid mutations. *Genome Res* **19**: 1562–1569.
- Kumar S, Dudley JT, Filipowski A, Liu L. 2011. Phylomedicine: An evolutionary telescope to explore and diagnose the universe of disease mutations. *Trends Genet* **27**: 377–386.
- Kunz S, Rojek JM, Kanagawa M, Spiropoulou CF, Barresi R, Campbell KP, Oldstone MB. 2005. Posttranslational modification of α -dystroglycan, the cellular receptor for arenaviruses, by the glycosyltransferase LARGE is critical for virus binding. *J Virol* **79**: 14282–14296.
- 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.
- Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, et al. 2007. The diploid genome sequence of an individual human. *PLoS Biol* **5**: e254. doi: 10.1371/journal.pbio.0050254.
- Li Y, Vincenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersson G, Cao H, Korneliussen T, et al. 2010. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* **42**: 969–972.
- Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ, White TJ, Sunyaev SR, Nielsen R, et al. 2008. Proportionally more deleterious genetic variation in European than in African populations. *Nature* **451**: 994–997.
- Lohmueller KE, Albrechtsen A, Li Y, Kim SY, Korneliussen T, Vincenbosch N, Tian G, Huerta-Sanchez E, Feder AF, Grarup N, et al. 2011. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genet* **7**: e1002326. doi: 10.1371/journal.pgen.1002326.
- López Herráez D, Bauchet M, Tang K, Theunert C, Pugach I, Li J, Nandineni MR, Gross A, Scholz M, Stoneking M. 2009. Genetic variation and recent positive selection in worldwide human populations: Evidence from nearly 1 million SNPs. *PLoS ONE* **4**: e7888. doi: 10.1371/journal.pone.0007888.
- Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. 2011. Clan genetics and the complex architecture of human disease. *Cell* **147**: 32–43.
- Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci (Suppl 1)* **104**: 8597–8604.
- MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, et al. 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**: 823–828.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461**: 747–753.
- Marth GT, Yu F, Indap AR, Garimella K, Gravel S, Leong WF, Tyler-Smith C, Bainbridge M, Blackwell T, Zheng-Bradley X, et al. 2011. The functional spectrum of low-frequency coding variation. *Genome Biol* **12**: R84. doi: 10.1186/gb-2011-12-9-r84.
- McCarthy MI, Hirschhorn JN. 2008. Genome-wide association studies: Potential next steps on a genetic journey. *Hum Mol Genet* **17**: R156–R165.
- McKenna JJ, Thoman EB, Anders TF, Sadeh A, Schechtman VL, Glotzbach SF. 1993. Infant-parent co-sleeping in an evolutionary perspective: Implications for understanding infant sleep development and the sudden infant death syndrome. *Sleep* **16**: 263–282.
- Mekel-Bobrov N, Gilbert SL, Evans PD, Vallender EJ, Anderson JR, Hudson RR, Tishkoff SA, Lahn BT. 2005. Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. *Science* **309**: 1720–1722.
- Miller MP, Kumar S. 2001. Understanding human disease mutations through the use of interspecific genetic variation. *Hum Mol Genet* **10**: 2319–2328.
- Montoya J, Lopez-Gallardo E, Diez-Sanchez C, Lopez-Perez MJ, Ruiz-Pesini E. 2009. 20 years of human mtDNA pathologic point mutations: Carefully reading the pathogenicity criteria. *Biochim Biophys Acta* **1787**: 476–483.
- Mu XJ, Lu ZJ, Kong Y, Lam HY, Gerstein MB. 2011. Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project. *Nucleic Acids Res* **39**: 7058–7076.
- Myles S, Davison D, Barrett J, Stoneking M, Timpson N. 2008. Worldwide population differentiation at disease-associated SNPs. *BMC Med Genomics* **1**: 22. doi: 10.1186/1755-8794-1-22.
- Nakagome S, Mano S, Kozłowski L, Bujnicki JM, Shibata H, Fukumaki Y, Kidd JR, Kidd KK, Kawamura S, Oota H. 2012. Crohn's disease risk alleles on the *NOD2* locus have been maintained by natural selection on standing variation. *Mol Biol Evol* **29**: 1569–1585.
- Nakajima T, Wooding S, Sakagami T, Emi M, Tokunaga K, Tamiya G, Ishigami T, Umemura S, Munkhbat B, Jin F, et al. 2004. Natural selection

- and population history in the human angiotensinogen gene (*AGT*): 736 complete *AGT* sequences in chromosomes from around the world. *Am J Hum Genet* **74**: 898–916.
- Neel JV. 1962. Diabetes mellitus: A “thrifty” genotype rendered detrimental by “progress”? *Am J Hum Genet* **14**: 353–362.
- Nei M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York.
- Nei M, Suzuki Y, Nozawa M. 2010. The neutral theory of molecular evolution in the genomic era. *Annu Rev Genomics Hum Genet* **11**: 265–289.
- Nesse RM, Williams GC. 1996. *Why we get sick: The new science of Darwinian medicine*. Vintage Books, New York.
- Nesse RM, Stearns SC, Omenn GS. 2006. Medicine needs evolution. *Science* **311**: 1071.
- Nesse RM, Bergstrom CT, Ellison PT, Flier JS, Gluckman P, Govindaraju DR, Niethammer D, Omenn GS, Perlman RL, Schwartz MD, et al. 2009. Evolution in health and medicine Sackler colloquium: Making evolutionary biology a basic science for medicine. *Proc Natl Acad Sci (Suppl 1)* **107**: 1800–1807.
- Ng PC, Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* **7**: 61–80.
- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, et al. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**: 272–276.
- Nielsen R, Hubisz MJ, Hellmann I, Torgerson D, Andres AM, Albrechtsen A, Gutenkunst R, Adams MD, Cargill M, Boyko A, et al. 2009. Darwinian and demographic forces affecting human protein coding genes. *Genome Res* **19**: 838–849.
- Novembre J, Di Rienzo A. 2009. Spatial patterns of variation due to natural selection in humans. *Nat Rev Genet* **10**: 745–755.
- Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Brittenham G, Looreesuwan S, Clark AG, Tokunaga K. 2004. Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. *Am J Hum Genet* **74**: 1198–1208.
- Ohta T. 1992. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst* **23**: 263–286.
- Oota H, Pakstis AJ, Bonne-Tamir B, Goldman D, Grigorenko E, Kajuna SL, Karoma NJ, Kungulilo S, Lu RB, Odunsi K, et al. 2004. The evolution and population genetics of the *ALDH2* locus: Random genetic drift, selection, and low levels of recombination. *Am Hum Genet* **68**: 93–109.
- Osier MV, Pakstis AJ, Soodyall H, Comas D, Goldman D, Odunsi A, Okonofua F, Parnas J, Schulz LO, Bertranpetit J, et al. 2002. A global perspective on genetic variation at the *ADH* genes reveals unusual patterns of linkage disequilibrium and diversity. *Am J Hum Genet* **71**: 84–99.
- Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, et al. 2009. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* **19**: 826–837.
- Prentice AM, Hennig BJ, Fulford AJ. 2008. Evolutionary origins of the obesity epidemic: Natural selection of thrifty genes or genetic drift following predation release? *Int J Obes (Lond)* **32**: 1607–1610.
- Pritchard JK. 2001. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* **69**: 124–137.
- Pritchard JK, Cox NJ. 2002. The allelic architecture of human disease genes: Common disease-common variant...or not? *Hum Mol Genet* **11**: 2417–2423.
- Pushkarev D, Neff NF, Quake SR. 2009. Single-molecule sequencing of an individual human genome. *Nat Biotechnol* **27**: 847–852.
- Ramensky V, Bork P, Sunyaev S. 2002. Human non-synonymous SNPs: Server and survey. *Nucleic Acids Res* **30**: 3894–3900.
- Rockman MV, Hahn MW, Soranzo N, Loisel DA, Goldstein DB, Wray GA. 2004. Positive selection on *MMP3* regulation has shaped heart disease risk. *Curr Biol* **14**: 1531–1539.
- Rockman MV, Hahn MW, Soranzo N, Zimprich F, Goldstein DB, Wray GA. 2005. Ancient and recent positive selection transformed opioid *cis*-regulation in humans. *PLoS Biol* **3**: e387. doi: 10.1371/journal.pbio.0030387.
- Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, Yi C, Kreuziger J, Baldi P, Wallace DC. 2007. An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res* **35**: D823–D828.
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**: 832–837.
- Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varily P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. *Science* **312**: 1614–1620.
- Sabeti PC, Varily P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**: 913–918.
- Saunders MA, Hammer MF, Nachman MW. 2002. Nucleotide variability at *G6pd* and the signature of malarial selection in humans. *Genetics* **162**: 1849–1861.
- Sawyer SA, Hartl DL. 1992. Population genetics of polymorphism and divergence. *Genetics* **132**: 1161–1176.
- Schaner P, Richards N, Wadhwa A, Aksentijevich I, Kastner D, Tucker P, Gumucio D. 2001. Episodic evolution of pyrin in primates: Human mutations recapitulate ancestral amino acid states. *Nat Genet* **27**: 318–321.
- Speakman JR. 2008. Thrifty genes for obesity, an attractive but flawed idea, and an alternative perspective: The ‘drifty gene’ hypothesis. *Int J Obes (Lond)* **32**: 1611–1617.
- Stajich JE, Hahn MW. 2005. Disentangling the effects of demography and selection in human history. *Mol Biol Evol* **22**: 63–73.
- Stearns SC. 1999. *Evolution in health and disease*. Oxford University Press, Oxford.
- Stearns SC, Nesse RM, Govindaraju DR, Ellison PT. 2010. Evolutionary perspectives on health and medicine. *Proc Natl Acad Sci (suppl 1)* **107**: 1691–1695.
- Stenson PD, Ball EV, Howells K, Phillips AD, Mort M, Cooper DN. 2009. The Human Gene Mutation Database: Providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum Genomics* **4**: 69–72.
- Subramanian S, Kumar S. 2006. Evolutionary anatomies of positions and types of disease-associated and neutral amino acid mutations in the human genome. *BMC Genomics* **7**: 306. doi: 10.1186/1471-2164-7-306.
- Talavera D, Taylor MS, Thornton JM. 2010. The (non)malignancy of cancerous amino acid substitutions. *Proteins* **78**: 518–529.
- Taylor RW, Turnbull DM. 2005. Mitochondrial DNA mutations in human disease. *Nat Rev Genet* **6**: 389–402.
- Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, Visscher PM. 2007. Recent human effective population size estimated from linkage disequilibrium. *Genome Res* **17**: 520–526.
- Teshima KM, Coop G, Przeworski M. 2006. How reliable are empirical genomic scans for selective sweeps? *Genome Res* **16**: 702–712.
- Thomas PD, Kejariwal A. 2004. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: Evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci* **101**: 15398–15403.
- Thompson EE, Kuttab-Boulos H, Witonsky D, Yang L, Roe BA, Di Rienzo A. 2004. *CYP3A* variation and the evolution of salt-sensitivity variants. *Am J Hum Genet* **75**: 1059–1069.
- Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drouiotou A, Dangerfield B, Lefranc G, Loiselet J, et al. 2001. Haplotype diversity and linkage disequilibrium at human *G6PD*: Recent origin of alleles that confer malarial resistance. *Science* **293**: 455–462.
- Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, et al. 2006. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* **39**: 31–40.
- Toomajian C, Kreitman M. 2002. Sequence variation and haplotype structure at the human *HFE* locus. *Genetics* **161**: 1609–1623.
- Toomajian C, Ajioka RS, Jorde LB, Kushner JP, Kreitman M. 2003. A method for detecting recent selection in the human genome from allele age estimates. *Genetics* **165**: 287–297.
- Travis JM, Munkemuller T, Burton OJ, Best A, Dytham C, Johnst K. 2007. Deleterious mutations can surf to high densities on the wave front of an expanding population. *Mol Biol Evol* **24**: 2334–2343.
- Trevathan W, Smith EO, McKenna JJ. 1999. *Evolutionary medicine*. Oxford University Press, New York.
- Tung J, Primus A, Bouley AJ, Severson TF, Alberts SC, Wray GA. 2009. Evolution of a malaria resistance gene in wild primates. *Nature* **460**: 388–391.
- Vander Molen J, Frisse LM, Fullerton SM, Qian Y, Del Bosque-Plata L, Hudson RR, Di Rienzo A. 2005. Population genetics of *CAPN10* and *GPR35*: Implications for the evolution of type 2 diabetes variants. *Am J Hum Genet* **76**: 548–560.
- Vasconcelos AT, Guimaraes AC, Castelletti CH, Caruso CS, Ribeiro C, Yokaiichiya F, Armoa GR, Pereira Gda S, da Silva IT, Schrago CG, et al. 2005. MamMiBase: A mitochondrial genome database for mammalian phylogenetic studies. *Bioinformatics* **21**: 2566–2567.
- Verrilli BC, Tishkoff SA. 2004. Signatures of selection and gene conversion associated with human color vision variation. *Am J Hum Genet* **75**: 363–375.
- Vitkup D, Sander C, Church GM. 2003. The amino-acid mutational spectrum of human genetic disease. *Genome Biol* **4**: R72. doi: 10.1186/gb-2003-4-11-r72.

- Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci* **102**: 18508–18513.
- Voight BF, Kudravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol* **4**: e72. doi: 10.1371/journal.pbio.0040072.
- Wall JD, Lohmueller KE, Plagnol V. 2009. Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Mol Biol Evol* **26**: 1823–1827.
- Wang E, Ding YC, Flodman P, Kidd JR, Kidd KK, Grady DL, Ryder OA, Spence MA, Swanson JM, Moyzis RK. 2004. The genetic architecture of selection at the human dopamine receptor D4 (*DRD4*) gene locus. *Am J Hum Genet* **74**: 931–944.
- Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Guo Y, et al. 2008. The diploid genome sequence of an Asian individual. *Nature* **456**: 60–65.
- Ward LD, Kellis M. 2012. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**: D930–D934.
- Welch J, Eyre-Walker A, Waxman D. 2008. Divergence and polymorphism under the nearly neutral theory of molecular evolution. *J Mol Evol* **67**: 418–426.
- Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, et al. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**: 872–876.
- Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R. 2007. Localizing recent adaptive evolution in the human genome. *PLoS Genet* **3**: e90. doi: 10.1371/journal.pgen.0030090.
- Wilmoth JR. 2000. Demography of longevity: Past, present, and future trends. *Exp Gerontol* **35**: 1111–1129.
- Wright A, Charlesworth B, Rudan I, Carothers A, Campbell H. 2003. A polygenic basis for late-onset disease. *Trends Genet* **19**: 97–106.
- Wu DD, Zhang YP. 2010. Positive selection drives population differentiation in the skeletal genes in modern humans. *Hum Mol Genet* **19**: 2341–2346.
- Yang Z. 2006. *Computational molecular evolution*. Oxford University Press, Oxford.
- Zheng D, Frankish A, Baertsch R, Kapranov P, Reymond A, Choo SW, Lu Y, Denoeud F, Antonarakis SE, Snyder M, et al. 2007. Pseudogenes in the ENCODE regions: Consensus annotation, analysis of transcription, and evolution. *Genome Res* **17**: 839–851.
- Zhu Q, Ge D, Maia JM, Zhu M, Petrovski S, Dickson SP, Heinzen EL, Shianna KV, Goldstein DB. 2011. A genome-wide comparison of the functional properties of rare and common genetic variants in humans. *Am J Hum Genet* **88**: 458–468.
- Zia A, Moses AM. 2011. Ranking insertion, deletion and nonsense mutations based on their effect on genetic information. *BMC Bioinformatics* **12**: 299. doi: 10.1186/1471-2105-12-299.



Human genomic disease variants: A neutral evolutionary explanation

Joel T. Dudley, Yuseob Kim, Li Liu, et al.

Genome Res. 2012 22: 1383-1394 originally published online June 4, 2012

Access the most recent version at doi:[10.1101/gr.133702.111](https://doi.org/10.1101/gr.133702.111)

References This article cites 159 articles, 57 of which can be accessed free at:
<http://genome.cshlp.org/content/22/8/1383.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>
