Identification of deleterious mutations within three human genomes

Sung Chun[1] and Justin C. Fay[1,2,3]

[1]*Computational Biology Program, Washington University, St. Louis, Missouri 63108, USA;* [2]*Department of Genetics, Washington University, St. Louis, Missouri 63108, USA*

Each human carries a large number of deleterious mutations. Together, these mutations make a significant contribution to human disease. Identification of deleterious mutations within individual genome sequences could substantially impact an individual's health through personalized prevention and treatment of disease. Yet, distinguishing deleterious mutations from the massive number of nonfunctional variants that occur within a single genome is a considerable challenge. Using a comparative genomics data set of 32 vertebrate species we show that a likelihood ratio test (LRT) can accurately identify a subset of deleterious mutations that disrupt highly conserved amino acids within protein-coding sequences, which are likely to be unconditionally deleterious. The LRT is also able to identify known human disease alleles and performs as well as two commonly used heuristic methods, SIFT and PolyPhen. Application of the LRT to three human genomes reveals 796–837 deleterious mutations per individual, ~40% of which are estimated to be at <5% allele frequency. However, the overlap between predictions made by the LRT, SIFT, and PolyPhen, is low; 76% of predictions are unique to one of the three methods, and only 5% of predictions are shared across all three methods. Our results indicate that only a small subset of deleterious mutations can be reliably identified, but that this subset provides the raw material for personalized medicine.

[Supplemental material is available online at http://www.genome.org.]

Mutations that impact an organism's ability to survive and reproduce are deleterious and must be eliminated by natural selection in order to ensure the long-term survival of a species (Lynch et al. 1993). Removal of deleterious mutations from the gene pool requires a substantial number of genetic deaths and incurs a considerable reproductive cost (Muller 1950). However, many deleterious mutations persist for hundreds of generations or more before they are removed, since their effects are largely masked in the heterozygous state (Simmons and Crow 1977).

The presence of deleterious mutations within the human population has a significant impact on human health. Inbreeding causes an increase in child morbidity and mortality and suggests that each human carries a sufficient number of deleterious mutations that, if made homozygous, would together result in premature death (Morton et al. 1956). In addition, most mutations that cause monogenic diseases are clearly deleterious. Diseases with a complex genetic basis are also likely to be affected by deleterious mutations. In support of this possibility, rare variants that have been associated with complex human diseases are often predicted to be deleterious (Cohen et al. 2004; Ahituv et al. 2007). However, even when rare variants have been associated with human disease, there is considerable uncertainty as to which rare variants are responsible for the association.

A number of methods have been developed to identify deleterious and/or disease-causing mutations within protein-coding sequences. These methods predict whether an amino acid altering mutation is deleterious or disease causing based on physicochemical properties (Grantham 1974), population frequency (Ohta 1973; Fay et al. 2001), protein structure (Chasman and Adams 2001; Sunyaev et al. 2001; Wang and Moult 2001), and cross-species conservation (Chasman and Adams 2001; Ng and Henikoff 2001; Sunyaev et al. 2001). For a comprehensive review of methods, see Ng and Henikoff (2006). While these methods can identify 40%–90% of disease-causing mutations, the rate of false-positives is uniformly high, 10%–20% (Ng and Henikoff 2006). The high rate of false-positives may be caused by most predictions relying on some aspect of sequence conservation, which is difficult to accurately model. One factor that confounds evolutionary models is that not all disease-causing mutations are conserved, presumably due to compensatory changes elsewhere in the protein (Kondrashov et al. 2002). A further complication is that mutations in highly conserved sequences do not always produce phenotypes that are easily noticeable (Ahituv et al. 2007; Hillenmeyer et al. 2008). Regardless of the cause, the accuracy of cross-species conservation depends on both the assumptions and parameterizations of evolutionary models that relate sequence conservation to fitness or function (Ng and Henikoff 2006; Care et al. 2007).

Genome sequencing of a large number of closely related species makes it possible to develop better parameterized evolutionary models that more accurately predict human deleterious mutations. Closely related species minimize the frequency of compensatory changes that enable functional sites to diverge. Conversely, a large number of species maximize the phylogenetic distance among taxa required to accurately distinguish selectively constrained sites from neutral sites that have not yet diverged (Eddy 2005). A number of models use putatively neutral classes of sites to distinguish functionally constrained and neutral sites based on closely related genomes (e.g., Stone et al. 2005; Asthana et al. 2007; Doniger et al. 2008). Although evolutionary models may not identify all disease-causing mutations, they provide a probabilistic framework in which the subset of disease-causing mutations that disrupt highly conserved amino acid positions can be accurately identified given enough phylogenetic information.

Prediction of deleterious mutations within individual human genomes has the potential to impact both the prevention and treatment of disease at an individual level. Although a number of

[3]Corresponding author.
E-mail jfay@genetics.wustl.edu; fax (314) 632-2156.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.092619.109.

human genomes have been sequenced, the number of nonsynonymous variants predicted to impact protein function varies widely. Using SIFT (Ng and Henikoff 2003), 14% (1455) of nonsynonymous variants within the Venter genome were predicted to impact protein function (Ng et al. 2008). Using PolyPhen (Ramensky et al. 2002), 7.3% (~770) of nonsynonymous variants within the Watson genome were predicted to impact protein function (Wheeler et al. 2008). However, comparison of these predictions is difficult since they are based on different data, different models, and use different methods to control for sequencing errors, a potentially important source of false-positives (Ng et al. 2008).

Table 1. Summary of deleterious mutations found in three individuals and the reference genome

		Tested		Deleterious	
Genome	High-quality variants	Number	Heterozygotes (percent)[a]	Number[b]	Heterozygotes (percent)[a]
J. Craig Venter	7534	5645	52	796 (14%)	78
James D. Watson	7353	5417	49	816 (15%)	76
Han Chinese	7462	5707	58	837 (15%)	83
Reference	NA	10,689	NA	838 (8%)	NA

[a]The frequency of heterozygotes was derived from genotype calls in the original publications.
[b]The percentage of tested mutations that are deleterious is shown in parentheses.
NA, Not available.

To identify and characterize deleterious mutations present within an individual genome, we examined three recently sequenced human genomes (Levy et al. 2007; Wang et al. 2008; Wheeler et al. 2008). Using a likelihood ratio test (LRT) (Doniger et al. 2008), we identified 796–837 amino acid altering mutations per genome that disrupt highly conserved amino acids. Comparison with two other methods, SIFT and PolyPhen, revealed only a small amount of overlap among the three methods and suggests that multiple methods should be used when trying to identify deleterious mutations in humans.

Results

Identification of deleterious mutations in three human genomes

Mutations that alter evolutionarily conserved sequences are likely to be deleterious and have a negative impact on fitness. To distinguish functionally constrained and unconstrained amino acid positions, we generated a comparative genomics data set using 32 vertebrate species (Methods). Multiple alignments of orthologous protein-coding sequences were generated for 18,993 human genes. The synonymous substitution rate was estimated to be 12.2 across all species and 4.7 across the eutherian clade, placental mammals, similar to previous studies (Methods). The large amount of divergence among these species implies that unconstrained amino acid positions should rarely be conserved across species, and mutations that alter conserved amino acids are likely to be deleterious.

To identify deleterious mutations present within an individual human genome, we examined nonsynonymous variants present within the genomes of J. Craig Venter, James D. Watson, and a Han Chinese male (Levy et al. 2007; Wang et al. 2008; Wheeler et al. 2008). To eliminate sequencing errors, we only used a subset of high-quality alleles, a *phred* quality score of 60 or greater for the Venter and Chinese genome, and a variant score of 70 or greater for the Watson genome (Wheeler et al. 2008). After eliminating 24%–26% of variants that occur in regions without a sufficient number of aligned orthologs for accurate inference of functional constraint, less than 10 eutherian species, we analyzed 5417–5707 nonsynonymous variants per individual.

Using a LRT, we identified between 796 and 837 deleterious mutations in the three diploid genomes, ~15% of those tested ($P < 0.001$; Table 1). The LRT compares the probability of the data under a conserved model, allowing for any level of selective constraint, relative to a neutral model, where there is no difference between the nonsynonymous and synonymous substitution rate (Methods). Most of the deleterious mutations, 59%–62%, were individual specific (Fig. 1). In addition, the majority of deleterious mutations, 76%–83%, were present in a heterozygous state. However, the true frequency of heterozygotes is likely higher, since homozygotes are more easily identified than heterozygotes, and some heterozygotes may be misclassified as homozygotes (Levy et al. 2007; Wang et al. 2008; Wheeler et al. 2008). In addition to deleterious mutations present within the three individual genomes, we identified another set of 838 deleterious mutations that occur in the reference genome in comparison to either the Venter, Watson, or Chinese genomes (Methods). Out of the 838 deleterious mutations, 474 were specific to the reference genome and were not present in the Venter, Watson, or Chinese genome.

Consistent with previous studies (Fay et al. 2002), the frequency of deleterious mutations is lower on the X chromosome compared with the autosomes, 10.5% relative to 18.1%, respectively ($P < 0.05$, Fisher's exact test). Only 14 out of 1928 deleterious mutations were found on the X relative to 119 out of 8761 neutral variants. These results indicate that partially recessive deleterious mutations are more rapidly eliminated from the X chromosome.

An abundance of common deleterious mutations

Most deleterious mutations are maintained at low-population frequencies due to negative selection (Kryukov et al. 2007). However, 435/1928 (23%) of the deleterious mutations are present in more than one of the three genomes, suggesting that at least some of the deleterious mutations may be common (Fig. 1). Consistent with negative selection, the fraction of nonsynonymous alleles

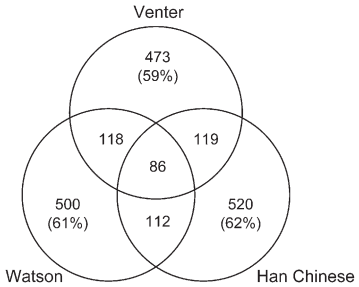


Figure 1. Venn diagram of deleterious mutations identified in J. Craig Venter, James D. Watson, and a Han Chinese individual. The percentage of individual-specific deleterious mutations found in each genome is shown in parentheses.

Table 2. Deleterious mutations are enriched in rare frequency classes

	Rare alleles		Common alleles[a]	
Genome	Tested	Deleterious[b]	Tested	Deleterious[b]
J. Craig Venter	766	213 (28%)	3066	583 (19%)
James D. Watson	940	303 (32%)	2632	513 (19%)
Han Chinese	703	186 (26%)	3438	651 (19%)

[a]Common alleles include those that are shared between any of the three genomes.
[b]Percent deleterious is shown in parentheses.

that are deleterious is lower for those that are shared among the three individuals, 19%, relative to those that are individual specific, 26%–32% (Table 2). To more accurately quantify the frequency of deleterious mutations we used the HapMap Phase II and Phase III panels and found that 1121/1928 deleterious mutations (58%) are common, greater than 5% allele frequency in at least one of the three HapMap panels. Surprisingly, many deleterious mutations have reached intermediate to high frequencies; 925 (48%) are at frequencies >20%, 472 (24.5%) are at frequencies >50%, and 163 (8.5%) are at frequencies >80%.

Deleterious mutations can become common if their effects are buffered by recently duplicated genes. Degeneration of recently duplicated genes is expected under both nonfunctionalization and subfunctionalization models of gene duplication (Lynch and Force 2000). Tabulating all deleterious mutations that occur in recently duplicated proteins (Methods), 70 out of 1928 deleterious mutations (3.6%) occur in duplicated genes (Fig. 2A). Although there is a significant enrichment of deleterious relative to neutral alleles in duplicated genes (χ^2, $P < 0.01$), only seven of the common deleterious alleles occur in duplicated genes. Thus, most common deleterious mutations are not buffered from selection by gene duplication.

If negative selection is weak, a substantial number of deleterious mutations can become common by random genetic drift. To examine whether common deleterious alleles are under weaker selection than those that are rare, we calculated the frequency of deleterious mutations in perfectly conserved sites, the frequency of deleterious mutations that caused radical amino acid substitutions, defined by a BLOSUM62 score of ≤ -2, and the frequency of deleterious mutations for which the deleterious allele was observed in one or more noneutherian species. Combining the data from all three genomes, rare deleterious mutations are more likely to occur in perfectly conserved sites, are more often radical, and are less likely to be present in noneutherian species (χ^2 test, $P < 0.01$; Fig. 2B). This result suggests that rare deleterious mutations are under stronger negative selection than common alleles, consistent with both empirical and theoretic studies (Keightley and Eyre-Walker 2007; Kryukov et al. 2007; Boyko et al. 2008). However, some of the common deleterious mutations may also be false-positives.

Estimation of the false-positive rate

A number of factors could lead to false-positive prediction of deleterious mutations. To estimate the rate of false-positives, we examined multiple hypothesis testing, sequencing errors, and model assumptions. Given an uncorrected *P*-value cutoff of 0.001 and the number of tests (Table 1), we estimate five to six false-positive predictions per individual due to multiple hypothesis testing. A *P*-value cutoff of 0.01 results in a prediction of 1120–1197 deleterious mutations within the three genomes, but is also expected to include a higher number of false-positives, 54–57 per individual.

Sequencing errors can also result in false-positive predictions. Given a *phred* quality score cutoff of 60, 0.47 false-positives are expected due to sequencing errors in the Venter genome (Methods). For the reference genome, 53 false-positives are expected, assuming three nucleotide substitution errors per 1 Mbp (Schmutz et al. 2004). The rate of sequencing errors is more difficult to know for the Watson and Chinese genomes, since a complex series of quality filters were used and quality values derived from new sequencing technologies may not be entirely accurate (e.g., Brockman et al. 2008).

To empirically estimate the impact of sequence errors on prediction of deleterious mutations, we tested whether the frequency of deleterious mutations is affected by the quality score cutoff. Consistent with a minor effect of sequencing errors, the fraction of deleterious mutations in the Venter genome is nearly constant for *phred* values greater than 60 (Supplemental Fig. S1A). High-quality nonsynonymous heterozygous variants from each genome were split into two groups of roughly equal numbers using their quality scores. In each case the proportion of variants called deleterious was slightly higher (0.3%–3.6%) in the group with higher compared with lower quality values (Supplemental Fig. S1).

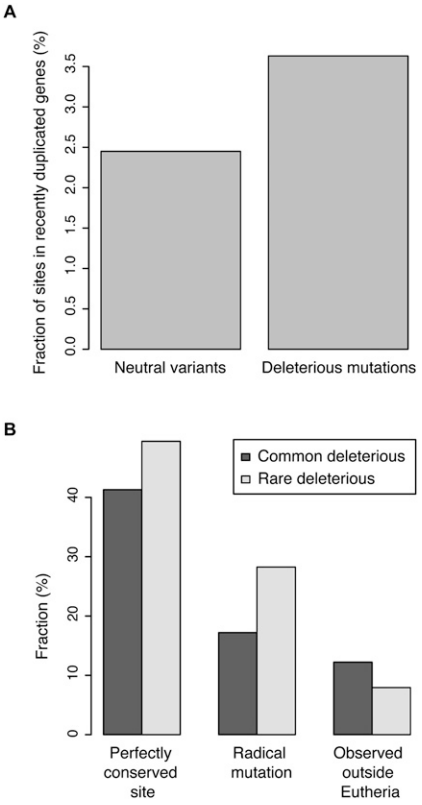


Figure 2. Characteristics of deleterious mutations. (*A*) Deleterious mutations ($n = 1928$) are more likely to occur in recently duplicated genes relative to neutral variants ($n = 8287$). (*B*) Mutations at perfectly conserved sites, mutations that cause radical amino acid changes, defined by BLOSUM62 ≤ -2, and mutations to amino acids that are not observed outside of eutherian mammals are more frequent among rare ($n = 807$) compared with common deleterious mutations ($n = 1121$).

This result suggests that few of the deleterious mutations identified by the likelihood ratio can be attributed to sequencing errors.

False-positive predictions can also arise if the assumptions of the LRT are violated. The LRT assumes a neutral substitution rate estimated from synonymous sites, and so false-positive predictions may occur for variants that occur in positions with lower than average mutation rates. Two significant sources of variation in mutation rates are methylated CpG sites and regional variation across the genome (Chimpanzee Sequencing and Analysis Consortium 2005; Gaffney and Keightley 2005). To examine the effect of regional variation across the genome, we estimated the number of deleterious mutations using a synonymous substitution rate of 10.1, two standard deviations below the mean. The standard deviation due to regional variation in mutation rates was obtained from the coefficient of variation (8.75%) reported for mouse–rat divergence in ancestral repeats at a scale of 1 Mb (Gaffney and Keightley 2005). To account for overestimation of the synonymous substitution rate due to methylated CpG sites, all CpG-prone sites were eliminated and the synonymous rate was found to be reduced by 12% in the mammalian clade and by 5% overall. Because regional variation in mutation rates generated a greater reduction in the synonymous rate, 1–10.1/12.2 (17%), we used a total synonymous rate of 10.1 and found 621 deleterious mutations remain significant in the Venter genome. This implies that only 22% (621/796) of highly conserved amino acid positions could be due to low mutation rates rather than negative selection. The true percentage is much lower since, only a small fraction of sites tested, ~2.5%, are expected to have a mutation rate two standard deviations below the genome average.

The false-positive rate can also be estimated using a set of negative controls. Previous studies estimated the rate of false-positives using nonsynonymous SNPs present in humans (Ng and Henikoff 2002) or substitutions between humans and other mammals (Ramensky et al. 2002). Given that a significant number of deleterious mutations and sequencing errors may be present within any sequenced genome, we generated a set of negative controls using 39,028 nonsynonymous substitutions that occurred after orangutan but before chimpanzee split off from the lineage leading to humans (Methods). The LRT identified 2633 substitutions (6.7%) as deleterious ($P < 0.001$). This is lower than that estimated for both PolyPhen (9%) (Sunyaev et al. 2001) and SIFT (20%) (Ng and Henikoff 2002), but is still much higher than that caused by sequencing errors or variation in mutation rate. When the negative controls were subdivided into three classes by the severity of amino acid changes using Grantham's distance (Grantham 1974), the LRT identified more conservative substitutions (8.3%) compared with moderate (6.1%) or radical substitutions (5.1%).

Estimation of the false-negative rate

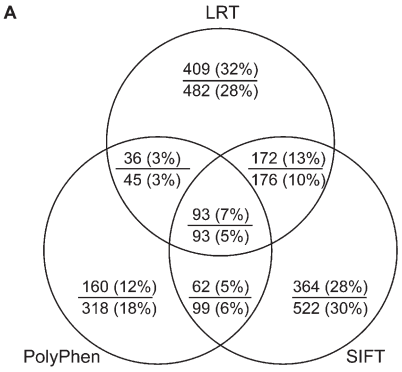
The false-negative rate was estimated using known disease-causing missense mutations in the OMIM database (Methods). Out of 5493 mutations, 3947 (71.9%) are significant by the LRT ($P < 0.001$). The false-negative rate of the LRT (28%) is similar to that of both SIFT (31%) (Ng and Henikoff 2002) and PolyPhen (31%) (Sunyaev et al. 2001). Of the 1546 disease mutations that were not identified, 644 (42%) occur at positions with little or no comparative genomic data (fewer than 10 eutherian mammals or a synonymous substitution rate ≤2). Although the LRT does not explicitly account for the type of amino acid change, the false-negative rate is lower for radical disease-causing mutations (25.2%) than for moderate or conservative disease-causing mutations (30.2% and 29.7%, respectively). Such differences are not due to varying availability of comparative genomic data. A similar fraction of mutations lack sufficient comparative genomic data in all three classes of amino acid changes.

Comparison to SIFT and PolyPhen predictions

A variety of methods have been developed to specifically identify mutations that cause human disease. Two of the most commonly used methods, SIFT (Ng and Henikoff 2003) and PolyPhen (Ramensky et al. 2002), use heuristic measures of cross-species conservation along with the type of amino acid change to predict human disease mutations. SIFT uses a median conservation score rather than synonymous sites to measure protein conservation. PolyPhen also uses a normalized cross-species conservation score and combines this with a variety of protein structural features when available. Both methods use nonredundant protein databases and so make use of a much more diverse set of species than the LRT.

Compared with the 796 deleterious mutations identified by the LRT, SIFT predicted 890 intolerable mutations and PolyPhen predicted 768 possibly damaging and 555 probably damaging mutations in the Venter genome (Methods). The overlap between these predictions is low but significantly greater than chance (Fig. 3). Out of all predictions, 18%, 30%, and 28% were specific to PolyPhen, SIFT, and the LRT, respectively, and 93 mutations (5%) were predicted by all three methods. The overlap of all three methods is greater than the 5.9 mutations (0.3%) expected by chance. Each of the methods predicted a similar fraction of

A



B

		LRT	
		Deleterious	Neutral
SIFT	Neutral	392 (7%)	4445 (79%)
	Deleterious	404 (7%)	404 (7%)

Figure 3. Comparison of SIFT, PolyPhen, and the likelihood ratio test (LRT) predictions. (*A*) Venn diagram of the number of predictions made by the three methods. Probably damaging mutations were used for PolyPhen. Numbers *below* and *above* each line are for the complete set of 7534 high-quality variants present within the Venter genome and a subset of 4303 where all three methods generated a prediction, respectively. (*B*) Overlap between the LRT and SIFT predictions based on the same alignments.

mutations that were not predicted by the other two methods; 57%, 59%, and 61% of predictions made by PolyPhen, SIFT, and the LRT, respectively, were not predicted by either of the two other methods.

The low overlap among predictions made by the LRT, SIFT, and PolyPhen is not due to differences in coverage. Out of the 7534 high-quality mutations present within the Venter genome, PolyPhen, SIFT, and the LRT generated predictions for 6746 (90%), 5401 (72%), and 5645 (75%) mutations, and all three methods generated predictions for 4303 (57%) mutations. Most of these differences are likely the result of each method requiring a sufficient number and diversity of aligned sequences in order to make a prediction, each method using a different set of sequences and alignments. Despite the differences in coverage, the overlap among the methods remains low when only the 4303 mutations with predictions made by all three methods were compared (Fig. 3).

To determine whether the low overlap can be attributed to alignment differences, SIFT was run using the same alignments as those used for the LRT. Figure 3B shows that tuning SIFT's median conservation score to generate a similar number of deleterious predictions as the LRT improved the overlap between SIFT and the LRT from 269 (19%) to 404 (34%). Thus, many, but not all of the differences between SIFT and the LRT are due to differences in the sequences and/or alignments used to identify evolutionary conserved amino acids.

The remaining differences between SIFT and the LRT predictions can be attributed to a number of factors. Out of 392 mutations that were predicted deleterious by the LRT but not SIFT, 226 (58%) occurred in highly conserved proteins and were not identified as intolerable by SIFT due to the high median conservation score of the protein. SIFT uses the median conservation score of a protein to eliminate predictions based on highly similar proteins. Because the LRT uses synonymous site divergence to calibrate conservation, many cases are likely false-negative predictions made by SIFT. In 184/392 (47%) cases the amino acid predicted to be deleterious by the LRT was found in one or more vertebrate species outside of the eutherian mammals. Although this indicates that these alleles are probably neutral in distantly related species, the LRT implies that amino acid conservation within the eutherian mammals is significant, and so the allele may be deleterious in humans.

A number of factors are also likely to contribute to the 404 mutations predicted to be intolerable by SIFT that were predicted neutral by the LRT. A total of 126/404 (31%) cases showed marginal significance by the LRT, ($0.001 < P < 0.01$). Some of these cases may be due to the lower power of the LRT, since SIFT differentiates between conservative and radical amino acid changes, a factor known to be predictive of function (Stone and Sidow 2005). The LRT may also have lower power for sites with a significant number of missing species. A total of 133/404 (33%) cases occurred in alignments for which there were not enough species to apply the LRT, as defined by a total synonymous substitution rate less than four. Because the LRT alignments are based mostly on closely related mammalian species, some of these cases are likely false-positive predictions made by SIFT. However, these predictions may also be a consequence of forcing SIFT to use alignments from closely related species and would not be predicted by SIFT when run using its own set of alignments.

Discussion

Identification of deleterious mutations within individual genomes has the potential to directly impact both health and reproductive decisions. The wealth of comparative genomics data now available makes it possible to rapidly identify all mutations that disrupt highly conserved amino acid positions that are likely to be deleterious. Here, we have evaluated the ability of a LRT to identify deleterious mutations within three human genomes. We identified a similar number of deleterious mutations, 796–837, in three human genomes, and showed that only a handful are expected to be false-positives due to sequencing errors or multiple hypothesis testing. While the LRT performs as well as two other commonly used methods, SIFT and PolyPhen, the overlap among predictions made by different methods is disturbingly low. Analysis of these differences indicates that both the algorithms, as well as the alignments used to identify conserved sites, make a significant contribution to the low overlap among predictions. Our results suggest that multiple methods should be used to reliably identify deleterious mutations for association with human disease.

The LRT is conceptually distinct from other comparative genomic methods. To our knowledge, all previous methods designed to identify deleterious mutations (for review, see Ng and Henikoff 2006) rely on heuristic procedures to distinguish sites within a protein that are conserved from those that are not conserved. This is achieved by selecting sequences that are not too closely or too distantly related to the sequence of interest and comparing the degree of conservation at the site of interest to other sites in the protein. The advantage of this approach is that the phylogenetic relationship and evolutionary distance among the sequences is not required. However, there are also limitations to this approach since no distinction is made between distantly related proteins that are highly conserved and closely related proteins that have evolved rapidly. Although proteins can always be selected that show a desired degree of similarity, compensatory changes are more likely to have occurred in distantly related proteins. In comparison, the LRT was designed to explicitly model phylogenetic relationships using a probabilistic framework and to minimize compensatory changes, which are thought to be common (Kondrashov et al. 2002) and can cause false-negatives by using closely related vertebrate species: mammals, chicken, frog, and fish.

The LRT also differs from other comparative genomic methods in that all amino acid changes are treated the same rather than weighting radical and conservative amino acid changes differently. While this is expected to reduce the power of the LRT, empirically, both the false-positive and the false-negative rates of the LRT are lower for radical relative to conservative amino acid changes. This may be a consequence of the genetic code and other mutational parameters being correlated with the ratio of radical to conservative amino acid changes (Dagan et al. 2002).

The LRT performs similar to two other commonly used comparative methods, SIFT (Sunyaev et al. 2001) and PolyPhen (Ng and Henikoff 2002). We found that the LRT was able to identify 72% of known disease-causing mutations, slightly higher than that reported for PolyPhen (69%) and SIFT (69%). However, it should be noted that these numbers are not directly comparable since accuracy depends on how it is measured. For example, SIFT and PolyPhen detected 70% and 72% of deleterious mutations when applied to the same protein mutation database (Bromberg and Rost 2007). Many, 644/1546 (42%), of the disease mutations that were not identified by the LRT can be attributed to the absence of sufficient comparative genomic data, i.e., not enough species. The remaining cases may be attributed to compensatory changes that allowed sites that cause human disease to change in non-human species.

The false-positive rate of the LRT is estimated to be lower than that estimated for SIFT and PolyPhen. While few false-positives are expected due to sequencing errors and multiple hypothesis testing, we found 6.7% of negative controls were called deleterious by the LRT. The rate of false-positives is lower than that estimated for both SIFT (20%–30%) and PolyPhen (9%–28%) (Sunyaev et al. 2001; Ng and Henikoff 2002; Bromberg and Rost 2007). However, some of these differences may be due to the use of different negative controls. Regardless of slight methodological differences, the frequency of putatively neutral mutations called deleterious (6.7%) is only half the fraction predicted to be deleterious in the Venter, Watson, and Chinese genomes (14%–15%), suggesting a false discovery rate of nearly 50% within individual genomes. Yet, the rate of false-positives estimated from the negative controls may be an overestimate, since not all human ancestral substitutions may be neutral. Some ancestral substitutions may be weakly deleterious mutations that became fixed by genetic drift, a process that can be exacerbated by a small effective population size. Other ancestral substitutions may be advantageous mutations that alter the function of previously conserved amino acids due to a change in the environment. Thus, ancestral substitutions provide an upper bound on the number of false-positive predictions.

The potentially high rate of false-positives may explain the large number of common alleles predicted to be deleterious in the three genomes. Similar to the negative controls, some of the common alleles may occur in sites that have been under negative selection in primates and other mammals but have recently become neutral along the human lineage. However, it is also possible that some deleterious mutations have increased in frequency due to hitchhiking along with recent positively selected mutations. Further work will be needed to address the impact of positive selection on the number and frequency of deleterious mutations.

Despite similar performance of the LRT, SIFT, and PolyPhen, the overlap among predictions made by the three methods is low, 5% (Fig. 3). The majority of differences are not due to cases where one or more methods did not make a prediction due to limited or insufficient data; the overlap remains low for cases where all three methods generated predictions (Fig. 3). Inspection of cases where the three methods disagreed revealed two general explanations for the disagreements. First, distantly related species can have a strong influence on the prediction of deleterious mutations and each method uses a different set of distantly related species. Distantly related species tend to have a large effect since fewer sites are conserved and they are more likely to carry the deleterious allele due to compensatory changes. The impact of distantly related species is significant, since each method measures conservation using a different set of distantly related species. Both SIFT and PolyPhen can use any homologous protein sequence and generate alignments using different nonredundant databases of protein sequences. In contrast, the LRT only uses a few nonmammalian species: chicken, frog, and five fish species (Supplemental Fig. S2). One of the goals of developing the LRT was to avoid using distantly related species that are more likely to contain compensatory changes, which may produce variable results depending on arbitrary decisions as to which distantly related species to use. However, the use of closely related species has its own set of disadvantages (see below).

A second class of disagreements are sites that are perfectly conserved within each of the different alignments but are slightly above or below cutoffs used by the LRT, SIFT, or PolyPhen. These borderline cases may also depend on which sequences are included in the alignment, because SIFT and PolyPhen use a site-specific score that is normalized to conservation within the rest of the protein and the *P*-value of the LRT explicitly depends on which species are included, since this determines the expected rate of change as measured by the synonymous substitution rate. One of the goals of developing the LRT was to accurately account for each species' contribution to the likelihood using the synonymous substitution rate. The drawback of this approach is that the increase in accuracy depends on a number of parameters that must be estimated from the data.

The LRT's use of closely related species may result in false-positive and false-negative predictions not made by either SIFT or PolyPhen. In order to use closely related species, the LRT uses synonymous sites to estimate the neutral substitution rate. However, the accuracy of this estimate depends on many factors. While we showed that CpG sites and large-scale regional variation in mutation rates are unlikely to have large effects on the prediction of deleterious mutations, other types of mutational variation were not accounted for (e.g., Hodgkinson et al. 2009). Even slight changes in the estimated neutral substitution rate will affect some predictions. Consequently, despite the advantages of using closely related species, many borderline cases may be false-positives or false-negatives. Borderline cases may be more accurately resolved by using additional closely related genomes or by inclusion of distantly species.

Is the power of the LRT limited by the amount of comparative genomic data? For perfectly conserved sites, the fraction of mutations called deleterious does not increase with the number of species used to identify conserved sites (Supplemental Fig. S3). However, the fraction of mutations predicted to be deleterious increases with the number of species used for sites that are not perfectly conserved (Supplemental Fig. S3). This suggests that additional species will only increase the power of the LRT to detect sites that are not perfectly conserved. Increasing the power to detect semi-conserved sites may be useful given the large number of human disease alleles that occur at sites that are not perfectly conserved.

What fraction of deleterious mutations was not identified? Not considering false-negatives due to compensatory changes and other modeling assumptions, deleterious mutations could be missed due to a lack of alignments or low power associated with the available alignments. To estimate the number of deleterious mutations that did not reach a *P*-value of <0.001, i.e., low power, we estimated the number of true positives as a function of the false discovery rate. Supplemental Figure S4 shows that as the *P*-value cutoff is lowered, the estimated number of true positives plateaus to ~1100 mutations, suggesting that nearly 75% of deleterious mutations were identified at a *P*-value cutoff of 0.001. The number of deleterious mutations missed due to a lack of sufficient alignments is more difficult to know. Only 74%–76% of mutations were tested in the Venter, Watson, and Chinese genomes due to a paucity of mammalian homologs. However, a smaller proportion of these may be deleterious, since they are more likely to occur in new, recently duplicated genes. An alternative method of estimating the number of deleterious mutations is based on allele frequency and avoids deficiencies due to a lack of alignments (Fay et al. 2001). Applying the allele frequency estimate to the Venter genome indicates that the LRT identified 62% of rare deleterious mutations (Methods). However, many factors other than the presence of suitable alignments may contribute to the difference between the allele frequency and the LRT estimates.

Despite different sequencing technologies, we found a very similar number of deleterious mutations within three human genomes, 796–837. Previous reports predicted ~770 (Wheeler

et al. 2008) and 1455 (Ng et al. 2008) missense mutations within the Watson and Venter genome, respectively. However, these differences can be almost entirely attributed to differences between the methods used to predict deleterious mutations and/or the quality score cutoffs used to eliminate sequencing errors. Using the same set of high-quality variants tested by the LRT, SIFT predicted 890, 769, and 861, and PolyPhen predicted 555, 501, and 496 deleterious mutations in the Venter, Watson, and Chinese genomes, respectively. While the coefficient of variation is greater than one for both SIFT and PolyPhen, it is less than one for the LRT. The standard deviation of the number of deleterious mutations identified by the LRT is 20.5 and is less than that expected assuming a Poisson distribution, 28.6. This supports the idea that truncating selection mediated by synergistic epistasis facilitates the removal of deleterious mutations (Crow and Kimura 1979). However, not all features of deleterious mutations were similar among the three genomes. The fraction of deleterious or neutral mutations within recently duplicated genes was much smaller in the Chinese compared with the other two genomes. This difference may be reflective of the greater difficulty of identifying SNPs in duplicated sequences using short-read sequencing technologies.

In conclusion, the abundance of mutations that disrupt highly conserved amino acid positions within three healthy human genomes implies that in most cases their phenotypic effects are either small or that the mutations are recessive. However, many of the mutations may produce large effects when homozygous. Although only a small number of mutations were predicted deleterious by the LRT, SIFT, and PolyPhen, the use of all three methods should provide an excellent source of candidates for association with human disease. Finally, since just over half of all deleterious mutations were found to be common, our results support the possibility that rare variants make a significant contribution to complex human diseases (Pritchard 2001; Kryukov et al. 2007).

Methods

Comparative genomic data set

Multiple sequence alignments of protein-coding sequences were generated from 32 vertebrate species (Supplemental Fig. S2). Orthologous protein sequences were downloaded from Ensembl (ftp://ftp.ensembl.org/pub/release-49/emf/ensembl_compara/homologies/), originally inferred by TreeBest (Vilella et al. 2009), aligned using MUSCLE (Edgar 2004), and then backtranslated into nucleotide alignments. All known selenocysteine residues were masked. After removing alignments with too few orthologous, <10 eutherian mammals, there were 18,993 alignments with an average of 16.5 species per alignment.

Likelihood ratio test (LRT)

The LRT was used to compare the null model that each codon is evolving neutrally, with no difference in the rate of nonsynonymous (d_N) to synonymous (d_S) substitution, to the alternative model that the codon has evolved under negative selection with a free parameter for the d_N/d_S ratio. The log-likelihood ratio (LLR) of the conserved relative to the neutral model is:

$$LLR = \log\frac{L(D|T,\theta,d_N=\hat{C}d_S)}{L(D|T,\theta,d_N=d_S)}$$

where D is an alignment of a single codon, T is a phylogenetic tree, d_N and d_S are the nonsynonymous and synonymous substitution rates of the codon, and $\hat{C}$ is the maximum likelihood estimate of d_N/d_S. The synonymous rate and θ, the parameters of the rate matrix, were estimated from the concatenated set of all codons without gaps, as described in the next paragraph. P-values were obtained by comparing twice the log likelihood ratio to a χ^2 distribution with one degree of freedom. The LRT was implemented using the MG94 codon model (Muse and Gaut 1994) combined with an HKY85 model (Hasegawa et al. 1985) to account for unequal base frequencies and differences in the rates of transitions and transversions. This was accomplished within a maximum likelihood framework using HyPhy (Pond et al. 2005).

The synonymous substitution rate was estimated from gap-free concatenated alignments of 1227 genes completely conserved across all 32 species, a total of 54 kb, using the same model as that described above. The estimated d_S values for the human to the mouse–rat ancestor (0.46), mouse lineage (0.10), and rat lineage (0.13) are similar to previous estimates, 0.457, 0.095, and 0.101, respectively (Cooper et al. 2005).

To make the predictions of the LRT available to the community for every potential nonsynonymous variant in the human genome, we applied it to 10,073,284 codons for which there were a sufficient number of aligned species. A total of 5,828,045 sites were significant ($P < 0.001$ and $d_N/d_S < 1$). Note that a substantial number of nonsynonymous variants at these positions may not be predicted deleterious if the mutant allele being tested is present within one or more of the eutherian mammals. The complete data set can be obtained by request from the investigators or downloaded from the corresponding investigator's website (http://www.genetics.wustl.edu/jflab/).

Identifying deleterious mutations

A complete catalog of SNPs were obtained for J. Craig Venter and a Han Chinese male from their respective websites (http://www.jcvi.org/cms/research/projects/huref/ and http://yh.genomics.org.cn), and for James D. Watson directly from Dr. David Wheeler (Baylor College of Medicine, Houston, TX). Nonsynonymous and synonymous SNPs were identified using known genes in Ensembl release 49. Coding SNPs in ambiguous reading frames, due to overlap of adjacent genes or frame shifts between known splice variants, or in known pseudogenes, were excluded.

To avoid sequencing errors, stringent quality filters were applied. Since quality scores were not available for each allele in the Venter genome, these were independently tabulated by mapping SNPs in the Venter genome to individual sequencing reads (ftp://ftp.ncbi.nih.gov/pub/TraceDB/Personal_Genomics/Venter/) and obtaining *phred* quality values for each allele from the sum of the quality scores supporting each allele. For each SNP, a 1000-bp flanking sequence around the SNP was extracted from the reference genome (NCBI build 36) and queried with MegaBlast against the Venter reads. Of the high-scoring blast hits (E-value $< 1 \times 10^{-100}$), only reads with perfect alignment across the 40 bp flanking each SNP were retained. Out of the original nonsynonymous SNPs, 22% failed to support a combined *phred* score of ≥60 and at least two supporting reads. By comparison to experimentally validated SNPs (Levy et al. 2007), this high-quality set of SNP has a lower rate of false-positive (0 out of 15) but a higher rate of false-negatives (4 out of 19) compared with the list of SNPs originally reported in Venter. The SNPs in the Chinese genome were also filtered using a *phred* quality cutoff of 60. SNPs in the Watson genome are associated with a variant score that is similar yet not identical to a *phred* quality score (Wheeler et al. 2008). Based on the distribution of quality variant scores, we used a quality variant cutoff of 70 for the Watson genome (Supplemental Fig. S1B).

Deleterious mutations were predicted by nonsynonymous SNPs that disrupt significantly constrained codons defined by the

LRT ($P < 0.001$) and a number of subsequent filters (Supplemental Table S1). First, positions with low power, <10 eutherian mammals, were eliminated. Second, a small number of sites with d_N significantly greater than d_S were discarded. Finally, positions where the derived deleterious allele occurred in another eutherian species were eliminated. Deleterious mutations were assigned to either the tested or reference genome depending on whether the reference or variant allele resulted in a lower d_N/d_S ratio. The total number of deleterious mutations includes all heterozygous or homozygous positions that differ from the reference genome except for homozygous positions where the reference allele rather than the variant allele was inferred to be deleterious. The number of deleterious mutations in the reference genome is the nonredundant set identified by comparison with the Venter, Watson, and Chinese genomes.

False-positive rate

The false-positive rate due to sequencing errors was estimated using *phred* quality scores. Using a *phred* quality cutoff of 60, 0.027 false-positive SNPs per 1 Mb are expected in the Venter genome. This calculation is based on an average *phred* quality score of 75.7. A total of 10.4 million codons were tested and 56.2% of these, 17.6 Mb, were estimated to be significantly constrained codons ($P < 0.001$), based on a random sample of 6357 codons. Using the estimated rate of sequencing errors and the total number of constrained codons, we expect a total of 0.47 false-positive SNPs due to sequencing errors. Using the same approach, 1.06 false-positives are estimated to occur in the Chinese genome. Using an error rate of three nucleotide substitutions per 1 Mbp for the reference genome (Schmutz et al. 2004), a total of 52.8 false-positives are expected.

To empirically estimate the rate of false-positives, the frequency of deleterious mutations was estimated for heterozygous sites with lower and higher quality values. Only heterozygous sites were used, since homozygous variants are less likely to be deleterious but are more likely to have high-quality values. For the Venter, Watson, and Chinese genome the quality value splits were 157, 129, and 97, respectively (Supplemental Fig. S1). In each case, the estimated number of false-positives was zero, since the fraction of deleterious mutations was higher using alleles with higher quality values.

The false-positive rate was also estimated using a set of negative controls. A negative control set was defined by ancestral amino acid substitutions that were inferred by maximum parsimony to have occurred along the lineage leading to humans after the split with the orangutan lineage but before the split with the chimpanzee lineage. We identified 39,028 amino acids that are the same between human and chimpanzee and between orangutan and macaque but differ between human and orangutan. The LRT was applied to the negative control set as if these ancestral substitutions were missense mutations present in the human population. The sequences of the primate species used to identify the set of negative controls (human, chimp, orangutan, and macaque) were excluded from the LRT. To examine different types of amino acid changes, the negative controls were subdivided into three classes by the severity of amino acid changes using the Grantham scale (Grantham 1974). Conservative amino acid changes were defined by a Grantham score of 50 or below, moderate changes by a Grantham score of 51 to 100, and radical changes by a score of greater than 100.

False-negative rate

The false-negative rate was estimated using disease-causing missense mutations in OMIM database. The coordinates of OMIM allelic variants were converted from protein-based residue positions to genomic coordinates. Each OMIM gene was mapped to a single representative RefSeq protein that provided the best match to the positions and amino acids of curated OMIM alleles. To eliminate potential mapping errors, we filtered out amino acid variants that are more than one mutation away from the mapped codon in the reference genome. Out of 9231 missense variants in OMIM, we mapped 5493 variants (59.5%) to the genome and tested each of them using the LRT. Identification of conservative, moderate, and radical amino acid changes was quantified using the same set of Grantham scores as that used for the analysis of false-positives.

Allele frequency

Filtered nonredundant allele frequencies in three HapMap analysis panels (CEU, CHB, and YRI) were downloaded from http://www.hapmap.org (Phase II+III, release 26) and used to identify SNPs with at least 5% frequency in the CEU, CHB, or YRI panels (The International HapMap Consortium 2007). Permission for the responsible use of the HapMap Phase III data was obtained from the International HapMap Consortium (D Altshuler and R Gibbs, pers. comm.).

Recent gene duplication

Recently duplicated genes were defined by human paralogs with >95% protein identity. Percent identity was calculated from human paralogs within the MUSCLE-generated multiple alignments of homologs from Ensembl. A total of 1232 human genes were identified as having at least one recently duplicated paralog in the human genome.

SIFT and PolyPhen predictions

SIFT 3.0 was downloaded and run locally to predict deleterious mutations in Venter using two different modes. First, SIFT was run using an independent set of alignments built using the TrEMBL 39.8 protein sequence database. Note that while TrEMBL contains many human sequence variants, SIFT eliminates highly similar sequences (Ng and Henikoff 2002). SIFT was able to generate alignments and predictions without errors for 6539 out of 7534 nonsynonymous variants. Filtering by the median conservation cutoff of 3.5 as in Ng et al. (2008), we obtained 5401 predictions and a total of 890 variants predicted to affect protein function. Second, SIFT was run with the same set of alignments used by the LRT. The median conservation score was set to 4.0 so that the number of SNPs predicted to impact protein function was similar to the number predicted by the LRT.

PolyPhen 1.15 was obtained from Shamil Sunyaev (Harvard Medical School) and run locally to generate predictions for the Venter genome. SWISS-PROT 56.8 and the latest version of the BLASTNR, PDB, and DSSP databases were used as input. PolyPhen generated predictions for 6746 nonsynonymous variants and called 768 possibly damaging and 555 probably damaging. Because PolyPhen does not generate allele-specific predictions, some of these predictions may be for Venter-reference differences, where the derived, deleterious allele is present in the reference but not Venter. Both SIFT and the LRT avoided this issue because they both generate allele-specific predictions.

Estimation of rare deleterious mutations

To estimate the number of deleterious mutations using allele frequencies, we compared the ratio of nonsynonymous (N) to synonymous (S) variants for rare versus common frequency classes (Fay et al. 2001). Because allele frequencies were not always available

and nonsynonymous alleles may be more often selected for genotyping, we used double hits in the dbSNP database as a proxy for rare versus common alleles, where double hits are defined by SNPs with at least two submissions. Within the three human genomes, the N/S ratio of double hit variants was 0.84–0.85 and nondouble-hit variants was 1.01–1.08, leading to an average estimate of 689 (18.6%) rare deleterious mutations. After accounting for the fact that only 48% of all deleterious mutations predicted by the LRT are present as double hits within dbSNP, and so do not contribute to the frequency calculation, we estimate that 62% of all rare deleterious mutations were identified by the LRT (0.52*816/689).

Acknowledgments

We thank members of the Fay laboratory and S. Sunyaev for helpful suggestions, S. Sunyaev for sending us the PolyPhen algorithm, and The International HapMap Consortium for making the Phase III genotyping data available. This work was supported by a National Institutes of Health grant (GM-080669) and an Alfred P. Sloan Research Fellowship to J.C.F.

References

Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, Hebert S, Doelle H, Ersoy B, Kryukov G, Schmidt S, et al. 2007. Medical sequencing at the extremes of human body mass. *Am J Hum Genet* **80:** 779–791.

Asthana S, Roytberg M, Stamatoyannopoulos J, Sunyaev S. 2007. Analysis of sequence conservation at nucleotide resolution. *PLoS Comput Biol* **3:** e254. doi: 10.1371/journal.pcbi.0030254.

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.

Brockman W, Alvarez P, Young S, Garber M, Giannoukos G, Lee WL, Russ C, Lander ES, Nusbaum C, Jaffe DB. 2008. Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res* **18:** 763–770.

Bromberg Y, Rost B. 2007. Snap: Predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res* **35:** 3823–3835.

Care MA, Needham CJ, Bulpitt AJ, Westhead DR. 2007. Deleterious SNP prediction: Be mindful of your training data! *Bioinformatics* **23:** 664–672.

Chasman D, Adams RM. 2001. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: Structure-based assessment of amino acid variation. *J Mol Biol* **307:** 683–706.

Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437:** 69–87.

Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. 2004. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305:** 869–872.

Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. 2005. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* **15:** 901–913.

Crow JF, Kimura M. 1979. Efficiency of truncation selection. *Proc Natl Acad Sci* **76:** 396–399.

Dagan T, Talmor Y, Graur D. 2002. Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. *Mol Biol Evol* **19:** 1022–1025.

Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang S, Fay JC. 2008. A catalog of neutral and deleterious polymorphism in yeast. *PLoS Genet* **4:** e1000183. doi: 10.1371/journal.pgen.1000183.

Eddy SR. 2005. A model of the statistical power of comparative genome sequence analysis. *PLoS Biol* **3:** e10. doi: 10.1371/journal.pbio.0030010.

Edgar RC. 2004. Muscle: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797.

Fay JC, Wyckoff GJ, Wu C-I. 2001. Positive and negative selection on the human genome. *Genetics* **158:** 1227–1234.

Fay JC, Wyckoff GJ, Wu C-I. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415:** 1024–1026.

Gaffney DJ, Keightley PD. 2005. The scale of mutational variation in the murid genome. *Genome Res* **15:** 1086–1094.

Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science* **185:** 862–864.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* **22:** 160–174.

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, et al. 2008. The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science* **320:** 362–365.

Hodgkinson A, Ladoukakis E, Eyre-Walker A. 2009. Cryptic variation in the human mutation rate. *PLoS Biol* **7:** e27. doi: 10.1371/journal.pbio.1000027.

The International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449:** 851–861.

Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* **177:** 2251–2261.

Kondrashov A, Sunyaev S, Kondrashov F. 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.

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.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154:** 459–473.

Lynch M, Bürger R, Butcher D, Gabriel W. 1993. The mutational meltdown in asexual populations. *J Hered* **84:** 339–344.

Morton NE, Crow JF, Muller HJ. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. *Proc Natl Acad Sci* **42:** 855–863.

Muller HJ. 1950. Our load of mutations. *Am J Hum Genet* **2:** 111–176.

Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* **11:** 715–724.

Ng P, Henikoff S. 2001. Predicting deleterious amino acid substitutions. *Genome Res* **11:** 863–874.

Ng P, Henikoff S. 2002. Accounting for human polymorphisms predicted to affect protein function. *Genome Res* **12:** 436–446.

Ng P, Henikoff S. 2003. Sift: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31:** 3812–3814.

Ng PC, Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* **7:** 61–80.

Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, Li K, Axelrod N, Busam DA, Strausberg RL, Venter JC. 2008. Genetic variation in an individual human exome. *PLoS Genet* **4:** e1000160. doi: 10.1371/journal.pgen.1000160.

Ohta T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* **246:** 96–98.

Pond S, Frost S, Muse S. 2005. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21:** 676–679.

Pritchard J. 2001. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* **69:** 124–137.

Ramensky V, Bork P, Sunyaev S. 2002. Human non-synonymous SNPs: Server and survey. *Nucleic Acids Res* **30:** 3894–3900.

Schmutz J, Wheeler J, Grimwood J, Dickson M, Yang J, Caoile C, Bajorek E, Black S, Chan YM, Denys M, et al. 2004. Quality assessment of the human genome sequence. *Nature* **429:** 365–368.

Simmons MJ, Crow JF. 1977. Mutations affecting fitness in *Drosophila* populations. *Annu Rev Genet* **11:** 49–78.

Stone E, Sidow A. 2005. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res* **15:** 978–986.

Stone EA, Cooper GM, Sidow A. 2005. Trade-offs in detecting evolutionarily constrained sequence by comparative genomics. *Annu Rev Genomics Hum Genet* **6:** 143–164.

Sunyaev S, Ramensky V, Koch I, Lathe W, Kondrashov A, Bork P. 2001. Prediction of deleterious human alleles. *Hum Mol Genet* **10:** 591–597.

Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. 2009. EnsemblCompara Genetrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19:** 327–335.

Wang Z, Moult J. 2001. SNPs, protein structure, and disease. *Hum Mutat* **17:** 263–270.

Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Zhang J, et al. 2008. The diploid genome sequence of an Asian individual. *Nature* **456:** 60–65.

Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen Y, Makhijani V, Roth GT, et al. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452:** 872–876.

Received February 10, 2009; accepted in revised form July 10, 2009.