PERSPECTIVE
https://doi.org/10.1038/s41588-019-0383-1

nature genetics

Measuring intolerance to mutation in human genetics

Zachary L. Fuller[1*], Jeremy J. Berg[1], Hakhamanesh Mostafavi[1], Guy Sella[1,2,3] and Molly Przeworski[1,2,3]

In numerous applications, from working with animal models to mapping the genetic basis of human disease susceptibility, knowing whether a single disrupting mutation in a gene is likely to be deleterious is useful. With this goal in mind, a number of measures have been developed to identify genes in which protein-truncating variants (PTVs), or other types of mutations, are absent or kept at very low frequency in large population samples—genes that appear 'intolerant' to mutation. One measure in particular, the probability of being loss-of-function intolerant (pLI), has been widely adopted. This measure was designed to classify genes into three categories, null, recessive and haploinsufficient, on the basis of the contrast between observed and expected numbers of PTVs. Such population-genetic approaches can be useful in many applications. As we clarify, however, they reflect the strength of selection acting on heterozygotes and not dominance or haploinsufficiency.

Experimental biologists and human geneticists are often interested in whether a single disrupting mutation, be it a PTV or a missense mutation, is likely to have a phenotypic effect[1–4]. A related question is whether such a mutation might lead to a decrease in the fitness of its carrier. The relationship between these two questions, between effects on phenotypes and on fitness, is not straightforward, and many potential paths exist from genotype to phenotype to fitness. For instance, a single mutation could lead to a severe clinical phenotype, thus indicating that the gene is haploinsufficient or that there is a gain of function, yet could have small or negligible effects on fitness unless homozygous. As examples, in *ELN* and *BRCA2*, a single PTV leads to a severe but late-onset disease, whereas homozygote PTVs are lethal[5–8]; thus, mutations in the genes are clearly haploinsufficient, but are they dominant with regard to fitness? Conversely, a mutation in a highly pleiotropic gene could have a very weak and potentially subclinical effect on any particular phenotype yet cumulatively inflict a severe cost on fitness[9].

Following common practice in human genetics, we refer to genes in which a single loss-of-function mutation has a discernible phenotypic effect in heterozygotes as 'haploinsufficient' (at least with regard to that phenotype)[4]. In turn, we describe genes in which a single disrupting mutation has an evolutionary fitness effect in heterozygotes as 'dominant' (Box 1). Although the term 'dominance' is also used to refer to the effect of a single allele on phenotype, for clarity, here, we restrict its use to denote effects on fitness. More precisely, following the convention in population genetics, we denote the fitnesses 1, $1 - hs$ and $1 - s$ as corresponding to genotypes AA, AD and DD, respectively, where D is the deleterious allele, h is the dominance coefficient, and s is the selection coefficient. Thus, a mutation is completely recessive if h is equal to 0—that is, if deleterious fitness effects are present only in homozygotes—and at least partially dominant otherwise.

Estimating the strength of selection acting on a gene in terms of the selection coefficient (s) and dominance effects (h) of mutations has a long history in population genetics[10–13]. In model organisms, such estimates have relied on mutation-accumulation experiments and assays of gene-deletion libraries[10,14–16]; in humans and other species, these parameters have been inferred from patterns of genetic variation[17–21]. The inferences are based on the notion of a mutation-selection-drift balance, namely that the frequencies of deleterious alleles in a sample reflect a balance between the rate at which they are introduced by mutation and the rate at which they are purged from the population by selection (as well as random changes in frequency due to genetic drift). Mutations with larger hs are purged more effectively and hence are expected to be present at lower frequencies in the population—or, equivalently, are more likely to be absent from large samples (Box 1). Therefore, one way to identify genes whose loss is likely to decrease fitness is to assess whether disrupting mutations are found at lower frequencies than expected under some sensible null model.

To our knowledge, the approach of prioritizing human disease genes on the basis of fitness consequences of disrupting mutations was introduced by Petrovski et al.[22], who ranked genes by comparing the observed number of common PTVs and missense mutations to the total number of observed variants. Their statistic was then supplemented by a number of others[23–26], notably pLI, which is defined as an estimate of the probability of being loss-of-function intolerant[27]. Loosely, pLI is derived from a comparison of the observed number of PTVs in a sample to the number expected in the absence of fitness effects (that is, under neutrality), given an estimated mutation rate for the gene. Specifically, Lek et al.[27] assumed that the number of PTVs observed in a gene is Poisson distributed with mean λM, where M is the number of segregating PTVs expected in a sample under neutrality (estimated for each gene according to a mutation model[23] and the observed synonymous polymorphism counts), and λ reflects the depletion in the number due to selection. The authors categorized genes as being neutral (with $\lambda_{\mathrm{Null}} = 1$), recessive ($\lambda_{\mathrm{Rec}} = 0.463$) or haploinsufficient ($\lambda_{\mathrm{HI}} = 0.089$). The fixed values of λ_{Rec} and λ_{HI} were obtained from the average proportional decrease in the number of observed PTVs in genes classified as recessive and severely haploinsufficient, respectively; the classification was based on phenotypic effects of mutations in the ClinGen dosage-sensitivity gene list and a hand-curated gene set of Mendelian

[1]Department of Biological Sciences, Columbia University, New York, NY, USA. [2]Department of Systems Biology, Columbia University, New York, NY, USA. [3]Program for Mathematical Genomics, Columbia University, New York, NY, USA. *e-mail: zlf2101@columbia.edu

Box 1 | Frequencies of deleterious alleles under mutation-selection-drift balance

Deleterious alleles are introduced into the population by mutation, then change in frequency as a result of the combined effects of genetic drift and natural selection. Unless a disease mutation confers an advantage in some environments (for example, the sickle-cell allele in populations with severe malaria[52]), the frequency at which it will be found in a population reflects a balance between the rate at which it is introduced by mutation and removed by purifying selection, modulated by the effects of genetic drift[42–44].

This phenomenon is referred to as mutation-selection-drift balance and is modeled as follows (example in ref. [53]). Let u be the mutation rate from the wild-type allele A to deleterious allele D. This mutation rate can be defined per site or per gene, by summing the mutation rate to deleterious alleles across sites (this simple summing is based on the implicit assumptions that there is no complementation and that compound heterozygotes for deleterious alleles have the same fitness effects as homozygotes[54]). The fitness of diploid individuals carrying genes with wild-type (A) or deleterious (D) alleles is given by:

Genotype:	AA	AD	DD
Fitness:	1	$1 - hs$	$1 - s$

where s is the selection coefficient, which measures the fitness of DD relative to AA, and h is the dominance coefficient, such that hs is the decrease in fitness of AD relative to AA. In population genetics, the term dominance (with respect to fitness) is often defined as $h > 0.5$. Here, however, we define a mutation as partially dominant as long as h is not near 0, because this criterion is directly relevant to the expected frequency of deleterious mutations[55].

In the limit of an infinite, panmictic population (that is, ignoring genetic drift and inbreeding), when $h > 0$ (and $hs >> u$), the equilibrium frequency of the deleterious allele (D), q, is approximately[43]:

$$q \approx u / hs$$

Notably, when $h > 0$, the equilibrium frequency q is determined by the strength of selection in heterozygotes (that is, hs, the joint effects of h and s), because deleterious homozygotes are too infrequent for selection on them to appreciably affect allele dynamics in the population. Hence, in this approximation, for a given hs, different combinations of h and s will yield the same frequency of q.

Under the same conditions, for a completely recessive allele ($h = 0$), q is well approximated by[43]:

$$q \approx \sqrt{\frac{u}{s}}$$

Here, the equilibrium frequency is determined by selection in homozygotes. In this limit of an infinite population size, the frequency corresponding to a recessive allele with a given $s > 0$ can also arise from a dominant allele for some value of $hs > 0$.

In a finite population, there is a distribution of deleterious-allele frequencies rather than a single (deterministic) value for any values of h and s. For a constant population size N, this distribution was derived by Wright[44] and is again a function of hs (assuming that $2Nhs >> 1$ and setting aside the case of sustained, high levels of inbreeding[56]). The resulting distribution can be highly variable, reflecting both stochasticity in the mutation process and the variance due to genetic drift. Dramatic changes in population size, as experienced by human populations, can also have a marked effect on the distribution of deleterious alleles. Regardless of these complications, distinguishing complete recessivity ($h = 0$) from small hs may still not be feasible, and, other than for complete recessivity, the expected allele frequency is a function of hs, not h and s separately[55].

disorders[28]. Given this model, Lek et al.[27] estimated the proportion of human genes in each of their three categories and then obtained the maximum a posteriori probability of any given gene belonging to each of the categories. Genes with a high probability (set at ≥0.9) of belonging to the haploinsufficient class were classified as 'extremely loss-of-function intolerant'[27].

pLI has been broadly used in human genetics to help identify genes in which a single disrupting mutation is likely to be clinically important[2,29–36]. It is also increasingly used in clinical annotation and in databases of mouse models, as an indicator of haploinsufficiency and dosage sensitivity[37–41]. In fact, however, pLI and related measures reflect only the strength of selection acting on heterozygotes and are not directly informative about dominance effects on fitness, let alone about the degree of haploinsufficiency with respect to a phenotype.

The reason can be understood in population-genetic terms: unless h is vanishingly small (or long-term inbreeding levels are very high), a decrease in the frequency of PTVs—and hence of PTV counts—is indicative of the strength of selection acting on heterozygotes, hs, and not of the two parameters h and s separately. This result derives from mutation-selection-drift balance theory developed by Haldane[42,43], Wright[44] and others[45] (Box 1). Intuitively, it reflects that, when fitness effects in heterozygotes are strong relative to genetic drift, deleterious alleles are kept at a low frequency in the population. Homozygotes for the deleterious allele are therefore exceedingly rare, and selection acts almost entirely through heterozygotes. As a result, the frequencies of PTVs in a sample—and therefore pLI and related measures—reflect the strength of selection acting on heterozygotes. This may be true even for those genes classified as phenotypically recessive by clinicians: although a much stronger phenotype is seen in homozygotes, a subtle fitness effect on heterozygotes can be sufficient to markedly decrease the frequency of disease mutations[46].

To illustrate this point, we used forward simulations to model how the observed counts of PTVs (and hence pLI) depend on h and s for a gene of typical length, considering both a constant-size population setting (Fig. 1a; details in legend) and a more realistic model for human demographic history[47] (Fig. 1b). As can be seen, markedly different combinations of h and s lead to indistinguishable distributions of PTV counts (and hence of pLI values), as long as hs is the same (Fig. 1a,b). More generally, the probability of observing a specific PTV count is maximized along a ridge corresponding to combinations of h and s that result in a given hs value (Fig. 1c). As a result, pLI can be near 1 even when the dominance coefficient h is small, provided that s is sufficiently large; pLI is therefore not indicative of dominance per se.

Although these considerations clearly indicate that pLI should be considered to reflect hs, it was not designed to be an estimator of this parameter, and it has several problematic features as such. First, for a given value of hs, the expected value of pLI varies with gene length (Fig. 2a). Second, for a typical gene length and a wide range of hs values (that is, ~10^{-3} to ~10^{-1}), the distribution of pLI is highly variable and bimodal, covering most of the range from 0 to 1 (Fig. 2b). Consequently, two genes with the same hs can be assigned

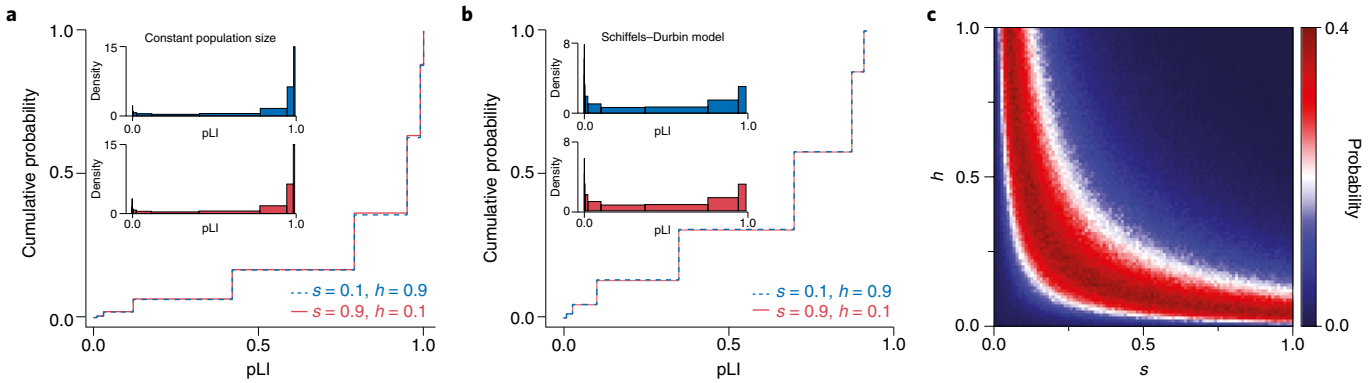

Fig. 1 | pLI relates to *hs*, but not *h* and *s* separately. a,**b**, Different combinations of *h* and *s* with the same *hs* value yield highly similar distributions of pLI. We considered PTVs arising in a hypothetical human gene of typical length for a population of constant size (**a**) and a plausible model of changes in the effective population size of Europeans over time (**b**)[47]. We modeled the distinct number of segregating PTVs in a population by using forward simulations (details in Supplementary Note). We first obtained the number of PTVs expected under neutrality by averaging more than 10^6 simulations with $s = h = 0$. Then, for different combinations of *s* and *h*, we calculated the pLI value for each replicate from the number of PTVs obtained. The lines show the cumulative distribution of pLI in 10^6 replicates for the parameter combinations of $s = 0.1$, $h = 0.9$ (blue, dashed) and $s = 0.9$, $h = 0.1$ (red, solid). The insets in each figure show the density of the distribution of pLI scores. **c**, The probability of observing a specific PTV count is maximized along a ridge of fixed *hs*. We generated the distribution of PTV counts in a hypothetical human gene under the same plausible demographic history as above (Schiffels–Durbin model) for a grid of *s* and *h* values, by using 10^6 replicates for each parameter combination. The figure depicts the likelihood of observing a PTV count of 3 (the value that by chance was obtained in the first run of $s = 0.10$, $h = 0.90$ and was treated as observed) for each combination of *h* and *s*.

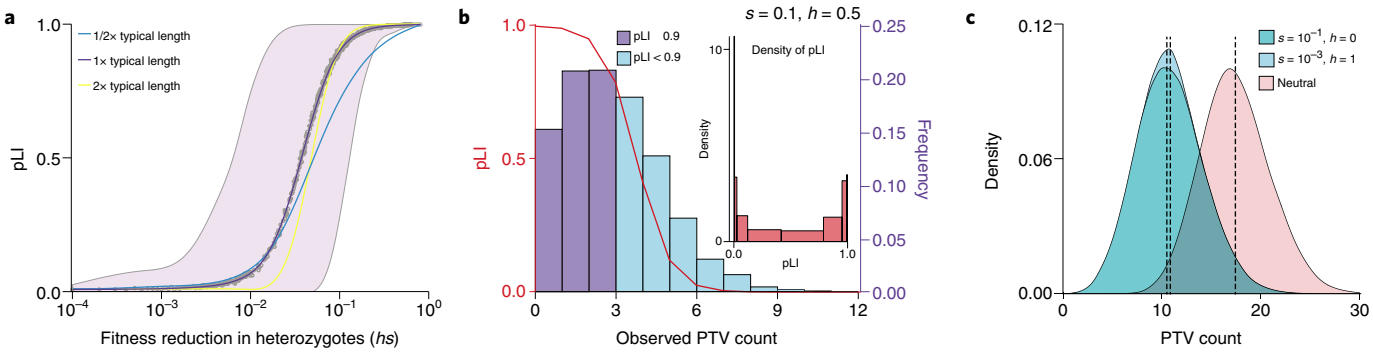

Fig. 2 | Properties of pLI. a, Behavior of pLI as a function of *hs*. We simulated the counts of PTVs for a range of *hs* values under a plausible model of population-size changes (Schiffels–Durbin model[47]; Supplementary Note). For each run, we calculated pLI by using the observed number of PTVs and the expected number obtained from averaging over neutral simulations. The purple line is the LOESS-smoothed curve over all simulations for each value of *hs* (the *x* axis on a $\log_{10}$ scale) in a human gene of typical length. The shaded area represents the central 95%-tile interval of pLI scores for each value of *hs*. The cyan and yellow lines are the LOESS-smoothed curves for simulations in a gene with half or twice the length of a typical gene, respectively. **b**, For a given *hs*, pLI scores are highly variable. The red curve depicts the pLI score as a function of the number of observed PTVs. The histogram represents the distribution of simulated PTV counts for $s = 0.1$, $h = 0.5$ under a plausible demographic model for Europeans[47], in a human gene of typical length; darker bars indicate scores that would be classified as extremely loss-of-function intolerant[27]. The inset shows the density of pLI scores. **c**, Complete recessivity ($h = 0$) and weak selection on heterozygotes ($hs > 0$) can lead to similar PTV counts. The distribution labeled 'neutral' shows the simulated counts of PTVs with *h* and *s* both equal to 0. Each distribution shows the results from 10^6 simulations. Dashed lines indicate the mean of each distribution.

radically different pLI values (Fig. 2b). Conversely, the same pLI value can reflect markedly different *hs* values, as illustrated by the large variance of pLI in the *hs* range between 10^{-3} and 10^{-1} (Fig. 2a). Outside this range of *hs* values, pLI is almost uninformative about the underlying parameter: below $hs \approx 10^{-3}$, pLI is ~0 for any value of *hs*; above $hs \approx 10^{-1}$, it is always ~1, and these properties worsen with increasing gene length (Fig. 2a). Our simulations further illustrate that for a given *hs*, genetic drift also contributes to the variance in PTV counts, a feature that is ignored in the construction of pLI (through its reliance on a Poisson distribution of PTV counts)[48]. Thus, if the goal is to learn about fitness effects to help prioritize disease genes, a direct estimate of *hs* (for example, those in refs. [48,49]) under a plausible demographic model, together with a measure of statistical uncertainty, would be preferable.

Recasting pLI in a population-genetic framework also aids in understanding why the assignment of genes as recessive is even less reliable (Fig. 2c). Lek et al.[27] aim to divide genes into three categories, two of which correspond to $hs > 0$ (pLI) and $hs = 0$, $s = 0$ (pNULL). Logically, the remaining category (pREC) should include completely recessive cases (that is, those with $hs = 0$ but $s > 0$), in which selection acts exclusively against homozygotes (Box 1). Regardless of the method used, however, distinguishing this category from the $hs > 0$ case is usually not feasible, because the same expected allele frequency (and hence PTV count) can arise when $h = 0$ or when $hs > 0$ but small (see Box 1 and Fig. 2c). For example, for a typical per-gene mutation rate to disease alleles of $u = 10^{-6}$ and no genetic drift, the frequency of disease alleles would be 1% whether $h = 0$ (completely recessive) and $s = 10^{-2}$ or $h = 1$ (fully dominant) and

$s = 10^{-4}$ (equations in Box 1). In other words, strongly deleterious, completely recessive PTVs can be difficult to distinguish from those that are weakly selected and at least partially dominant.

Why, then, in practice, are genes classified by clinicians as dominant on the basis of Mendelian disease phenotypes enriched in higher pLI scores when those classified as recessive are not[2,27,31]? Mendelian disease genes consist mostly of cases in which mutations are known to cause a highly deleterious outcome, that is, for which there is prior knowledge that s is likely to be large (even close to 1). When s is large, a gene will be classified by pLI as haploinsufficient as long as fitness effects in heterozygotes are sufficient to decrease the number of observed PTVs, that is, as long as h is not tiny. For most genes, however, there is no prior knowledge about s, and in that case, pLI—or any measure based on the frequency of PTVs—cannot reliably distinguish recessivity from dominance, let alone identify haploinsufficiency.

In summary, population-genetic approaches based on the deficiency of putatively deleterious mutations[2,3,23,25,49–51] hold great promise for prioritizing genes in which mutations are likely to be harmful in heterozygotes[22,49]. Recasting these approaches in terms of underlying population-genetic parameters provides a natural framework for their interpretation and a clearer understanding of what inferences they can reliably support: these approaches identify genes in which single PTVs are likely to have large fitness effects in heterozygotes. For this subset of genes, there is information about dominance when s is known a priori to be large and not otherwise. For any gene, the methods cannot be used to directly infer haploinsufficiency status. Where fitness effects are to be used as an indication of pathogenicity, we therefore argue that a better approach is the development of direct estimates of hs (and measures of uncertainty) under realistic demographic models for the population of interest.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

C++ source code for the simulations of PTV counts and accompanying scripts used for plotting and data analysis are available at https://github.com/zfuller5280/MutationIntoleranceSimulations.

Received: 31 July 2018; Accepted: 22 February 2019;
Published online: 8 April 2019

References

1. Blake, J. A., Bult, C. J., Eppig, J. T., Kadin, J. A. & Richardson, J. E. The Mouse Genome Database genotypes:phenotypes. *Nucleic Acids Res.* **37**, D712–D719 (2009).
2. Bartha, I., di Iulio, J., Venter, J. C. & Telenti, A. Human gene essentiality. *Nat. Rev. Genet.* **19**, 51–62 (2018).
3. Eilbeck, K., Quinlan, A. & Yandell, M. Settling the score: variant prioritization and Mendelian disease. *Nat. Rev. Genet.* **18**, 599–612 (2017).
4. Huang, N., Lee, I., Marcotte, E. M. & Hurles, M. E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* **6**, e1001154 (2010).
5. Raybould, M. C., Birley, A. J. & Hultén, M. Molecular variation of the human elastin (ELN) gene in a normal human population. *Ann. Hum. Genet.* **59**, 149–161 (1995).
6. Wooster, R. et al. Identification of the breast cancer susceptibility gene *BRCA2*. *Nature* **378**, 789–792 (1995).
7. Wagenseil, J. E. et al. The importance of elastin to aortic development in mice. *Am. J. Physiol. Heart Circ. Physiol.* **299**, H257–H264 (2010).
8. Roy, R., Chun, J. & Powell, S. N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat. Rev. Cancer* **12**, 68–78 (2011).
9. Simons, Y. B., Bullaughey, K., Hudson, R. R. & Sella, G. A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biol.* **16**, e2002985 (2018).
10. Simmons, M. J. & Crow, J. F. Mutations affecting fitness in *Drosophila* populations. *Annu. Rev. Genet.* **11**, 49–78 (1977).
11. Keightley, P. D. The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics* **138**, 1315–1322 (1994).
12. Deng, H. W. & Lynch, M. Estimation of deleterious-mutation parameters in natural populations. *Genetics* **144**, 349–360 (1996).
13. Orr, H. A. Fitness and its role in evolutionary genetics. *Nat. Rev. Genet.* **10**, 531–539 (2009).
14. Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 335–355 (1972).
15. Phadnis, N. & Fry, J. D. Widespread correlations between dominance and homozygous effects of mutations: implications for theories of dominance. *Genetics* **171**, 385–392 (2005).
16. Agrawal, A. F. & Whitlock, M. C. Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* **187**, 553–566 (2011).
17. Williamson, S. H. et al. Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc. Natl Acad. Sci. USA* **102**, 7882–7887 (2005).
18. Eyre-Walker, A., Woolfit, M. & Phelps, T. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* **173**, 891–900 (2006).
19. Boyko, A. R. et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* **4**, e1000083 (2008).
20. Racimo, F. & Schraiber, J. G. Approximation to the distribution of fitness effects across functional categories in human segregating polymorphisms. *PLoS Genet.* **10**, e1004697 (2014).
21. Kim, B. Y., Huber, C. D. & Lohmueller, K. E. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics* **206**, 345–361 (2017).
22. Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S. & Goldstein, D. B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* **9**, e1003709 (2013).
23. Samocha, K. E. et al. A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).
24. Steinberg, J., Honti, F., Meader, S. & Webber, C. Haploinsufficiency predictions without study bias. *Nucleic Acids Res.* **43**, e101 (2015).
25. Bartha, I. et al. The characteristics of heterozygous protein truncating variants in the human genome. *PLOS Comput. Biol.* **11**, e1004647 (2015).
26. Fadista, J., Oskolkov, N., Hansson, O. & Groop, L. LoFtool: a gene intolerance score based on loss-of-function variants in 60 706 individuals. *Bioinformatics* **33**, 471–474 (2017).
27. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
28. Blekhman, R. et al. Natural selection on genes that underlie human disease susceptibility. *Curr. Biol.* **18**, 883–889 (2008).
29. Lelieveld, S. H. et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* **19**, 1194–1196 (2016).
30. Ruderfer, D. M. et al. Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. *Nat. Genet.* **48**, 1107–1111 (2016).
31. Kosmicki, J. A. et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Genet.* **49**, 504–510 (2017).
32. Skraban, C. M. et al. WDR26 haploinsufficiency causes a recognizable syndrome of intellectual disability, seizures, abnormal gait, and distinctive facial features. *Am. J. Hum. Genet.* **101**, 139–148 (2017).
33. Stankiewicz, P. et al. Haploinsufficiency of the chromatin remodeler BPTF causes syndromic developmental and speech delay, postnatal microcephaly, and dysmorphic features. *Am. J. Hum. Genet.* **101**, 503–515 (2017).
34. Nguyen, H. T. et al. Integrated Bayesian analysis of rare exonic variants to identify risk genes for schizophrenia and neurodevelopmental disorders. *Genome Med.* **9**, 114 (2017).
35. Zarrei, M. et al. De novo and rare inherited copy-number variations in the hemiplegic form of cerebral palsy. *Genet. Med.* **20**, 172–180 (2018).
36. Heyne, H. O. et al. De novo variants in neurodevelopmental disorders with epilepsy. *Nat. Genet.* **50**, 1048–1053 (2018).
37. Zech, M. et al. Haploinsufficiency of KMT2B, encoding the lysine-specific histone methyltransferase 2b, results in early-onset generalized dystonia. *Am. J. Hum. Genet.* **99**, 1377–1387 (2016).
38. Haller, M., Mo, Q., Imamoto, A. & Lamb, D. J. Murine model indicates 22q11.2 signaling adaptor *CRKL* is a dosage-sensitive regulator of genitourinary development. *Proc. Natl Acad. Sci. USA* **114**, 4981–4986 (2017).
39. Wang, J. et al. MARRVEL: integration of human and model organism genetic resources to facilitate functional annotation of the human genome. *Am. J. Hum. Genet.* **100**, 843–853 (2017).
40. Afzali, B. et al. BACH2 immunodeficiency illustrates an association between super-enhancers and haploinsufficiency. *Nat. Immunol.* **18**, 813–823 (2017).

41. Gosalia, N., Economides, A. N., Dewey, F. E. & Balasubramanian, S. MAPPIN: a method for annotating, predicting pathogenicity and mode of inheritance for nonsynonymous variants. *Nucleic Acids Res.* **45**, 10393–10402 (2017).
42. Haldane, J. B. S. A mathematical theory of natural and artificial selection, part V: selection and mutation. *Math. Proc. Camb. Philos. Soc.* **23**, 838–844 (1927).
43. Haldane, J. B. S. The effect of variation of fitness. *Am. Nat.* **71**, 337–349 (1937).
44. Wright, S. The distribution of gene frequencies in populations. *Proc. Natl Acad. Sci. USA* **23**, 307–320 (1937).
45. Crow, J.F. & Kimura, M. *An Introduction to Population Genetics Theory* (Harper & Row, 1970).
46. Amorim, C. E. G. et al. The population genetics of human disease: the case of recessive, lethal mutations. *PLoS Genet.* **13**, e1006915 (2017).
47. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
48. Weghorn, D. et al. Applicability of the mutation-selection balance model to population genetics of heterozygous protein-truncating variants in humans. Preprint at https://www.biorxiv.org/content/10.1101/433961v1 (2018).
49. Cassa, C. A. et al. Estimating the selective effects of heterozygous protein-truncating variants from human exome data. *Nat. Genet.* **49**, 806–810 (2017).
50. Samocha, K.E. et al. Regional missense constraint improves variant deleteriousness prediction. Preprint at https://www.biorxiv.org/content/10.1101/148353v1 (2017).
51. Havrilla, J. M., Pedersen, B. S., Layer, R. M. & Quinlan, A. R. A map of constrained coding regions in the human genome. *Nat. Genet.* **51**, 88–95 (2018).
52. Piel, F. B. et al. Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nat. Commun.* **1**, 104 (2010).
53. Gillespie, J.H. *Population Genetics: a Concise Guide* (JHU Press, 2004).
54. Clark, A. G. Mutation-selection balance with multiple alleles. *Genetica* **102-103**, 41–47 (1998).
55. Simons, Y. B., Turchin, M. C., Pritchard, J. K. & Sella, G. The deleterious mutation load is insensitive to recent population history. *Nat. Genet.* **46**, 220–224 (2014).
56. Charlesworth, B. & Charlesworth, D. *Elements of Evolutionary Genetics* (W. H. Freeman, 2010).

Acknowledgements

We thank A. Chakravarti, G. Coop, M. B. Eisen, M. Hurles, J. K. Pritchard, Y. Shen and members of the laboratories of M. Przeworski and G. Sella for helpful discussions. This work was supported by GM128318 to Z.L.F., GM126787 to J.J.B., GM121372 to M.P. and GM115889 to G.S. We acknowledge computing resources from Columbia University's Shared Research Computing Facility project, which is supported by NIH Research Facility Improvement Grant 1G20RR030893-01, and associated funds from the New York State Empire State Development, Division of Science Technology and Innovation (NYSTAR) contract C090171.

Author contributions

All authors conceived and designed the project. M.P. and G.S. supervised the study. Z.L.F. performed simulations. H.M., J.J.B. and Z.L.F. led the data analysis. All authors wrote the manuscript and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0383-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to Z.L.F.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Zachary L. Fuller, Molly Przeworski

Last updated by author(s): Dec 18, 2018

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Custom c++ code was used to generate the simulations. Source code is available at: https://github.com/zfuller5280/MutationIntoleranceSimulations/PTV_count_simulations.cpp
Data analysis	Custom R and Python scripts were used to analyze the data. The custom functions used in analyses are available at https://github.com/zfuller5280/MutationIntoleranceSimulations/. All statistical analyses were conducted in R version 3.5.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

c++ source code for the simulations of PTV counts and accompanying scripts used for plotting and data analysis are available at https://github.com/zfuller5280/MutationIntoleranceSimulations.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size used in simulations is given by the sample size in the ExAC dataset and used in Lek et al. (2016).
Data exclusions	No data were excluded in the study.
Replication	No experimental data were collected. All simulations are reproducible and documented.
Randomization	Not applicable. There were not different experimental and control groups.
Blinding	Blinding is not relevant to this study. No grouping was performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging