Supplementary material for:

Estimation of allele-specific fitness effects across human protein-coding sequences and implications for disease

Yi-Fei Huang[1] and Adam Siepel[1]

[1]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

Details of the LASSIE model

Poisson Random Field model. LASSIE employs the Poisson Random Field (PRF) model to calculate the probability distribution of derived allele frequencies in the presence of natural selection, genetic drift, and new mutations[1,2]. These calculations rely on an infinite sites assumption, which implies at most one mutation per site on the time scale of human population genetics[3], thereby avoiding the complication of modeling multiple derived alleles per site. (Any sites with more than two alleles are removed from the input data; these sites are rare in the 1000 Genomes Project Yoruba data set.)

Let $f(y|S, \psi, \theta, t)$ be the probability density function for the derived allele frequency y at time t. Time is defined over the interval from $t = 0$, representing a deep ancestral population (prior to the emergence of any segregating polymorphisms), to the present, $t = t_{\text{current}}$. S represents the population-scaled selection coefficient, $S = 2N(0)s$, where $N(0)$ is the effective population at $t = 0$ and s is the genic selection coefficient associated with the derived mutation in question (see ref. 2). The genic selection model assumes that the fitnesses of heterozygous and homozygous carriers of the derived allele are equal to $1 + s$ and $1 + 2s$, respectively, relative to a fitness of 1 for homozygous carriers of the ancestral allele. ψ is a vector of demographic parameters that defines the relative effective

population size at time t, which is denoted $\rho(t, \psi)$ (as detailed below). Finally, $\theta = 4N(0)\mu$ is the population-scaled mutation rate, where μ is the mutation rate per generation per nucleotide site.

To ease numerical computation, we additionally apply a transformation of the allele frequencies, $g(y|S, \psi, \theta, t) = y(1-y)f(y|S, \psi, \theta, t)$. As demonstrated by Evans *et al.* (ref. 2), $g(y|S, \psi, \theta, t)$ can therefore be calculated by solving the partial differential equation,

$$\frac{\partial}{\partial t}g(y|S,\psi,\theta,t) = \underbrace{-Sy(1-y)}_{\text{strength of selection}}\ \frac{\partial}{\partial y}g(y|S,\psi,\theta,t) + \underbrace{\frac{y(1-y)}{2\rho(t,\psi)}}_{\text{strength of drift}}\ \frac{\partial^2}{\partial y^2}g(y|S,\psi,\theta,t), \quad (1)$$

with boundary conditions $\lim_{y\downarrow 0} = \theta\rho(t, \psi)$ and $\lim_{y\uparrow 1} = 0$.

Equation 1 describes how allele frequencies change stochastically over time, in response to the population genetic parameters associated with natural selection (S), genetic drift (ψ), and mutation (θ). To calculate the distribution of allele frequencies in the modern population, $f(y|S, \psi, \theta, t_{\text{current}})$, LASSIE solves Equation 1 numerically using the Crank-Nicolson algorithm[4]. More specifically, LASSIE discretizes the transformed allele frequency $g(y|S, \psi, \theta, t)$ into 1000 equal-size bins to form a discrete approximation of allele-frequencies. Similarly, it discretizes the scaled time t with a bin size of 0.0001. Then, it applies the Crank-Nicolson algorithm to solve Equation 1 iteratively, forward over time, to calculate $g(y|S, \psi, \theta, t_{\text{current}})$. Finally, the discretized density function of allele frequency in the modern population, $f(y|S, \psi, \theta, t_{\text{current}})$, is calculated using the inverse transformation, $f(y|S, \psi, \theta, t_{\text{current}}) = \frac{g(y|S,\psi,\theta,t_{\text{current}})}{y(1-y)}$.

Sampling distribution of derived allele frequencies. The function $f(y|S, \psi, \theta, t_{\text{current}})$ represents the population-level distribution of derived allele frequencies, but in practice, we can only obtain a finite number of samples from the population. Therefore, it is essential to specify the sampling distribution of derived allele frequencies. Let M_i be the (haploid) sample size at site i. For example, the high-coverage Yoruba data set used in this study consists of 51 unrelated individuals, so $M_i = 102$ for all the autosomal sites

without missing data. (Missing data is naturally handled in this framework by setting M_i equal to the number of alleles actually available at site i.) The probability of observing a polymorphic site i with m_i copies of the derived allele is simply given by the expectation of the binomial sampling distribution with respect to the density of continuous population-level allele frequencies[2],

$$Q(m_i|S,\psi,\theta) = \int_0^1 \underbrace{\binom{M_i}{m_i} y^{m_i}(1-y)^{M_i-m_i}}_{\text{binomial distribution}} \underbrace{f(y|S,\psi,\theta,t_{\text{current}})}_{\substack{\text{population-level} \\ \text{allele frequency}}} dy, \quad 1 \leq m_i \leq M_i - 1. \quad (2)$$

LASSIE employs Equation 4.1.18 in *Numerical Recipes in C* (ref. 5) to numerically calculate the integral $Q(m_i|S,\psi,\theta)$, using the discretization scheme defined for the Crank-Nicolson algorithm (above).

Equation 2 assumes that the ancestral allele is known. In practice, we consider uncertainty in the reconstructed ancestral allele, as previously estimated in a phylogenetic analysis[6,7] (see Online Methods). In particular, let q_i be the probability that the reconstructed ancestral allele is identical to the reference allele, and let m_i be the number of observed alternative (non-reference) alleles. Then the sampling distribution of segregating mutations can be calculated as,

$$P(m_i|S,\psi,\theta) = q_i Q(m_i|S,\psi,\theta) + (1-q_i)Q(M_i - m_i|S,\psi,\theta). \quad (3)$$

Note that, for simplicity and speed, we abandon explicit handling of uncertainty in ancestral alleles when using the model in the context of the mixture density network (see below).

Demographic model. To control the effect of population expansions on the distribution of allele frequencies, we employ a three-epoch demographic model in which $\rho(t,\psi)$ is a step function with two change points. Previous studies suggest that a simple demographic model with two to three epochs is powerful enough to account for the impact of human expansions on the site frequency spectrum[8–10]. In our three-epoch model, the

vector of demographic parameters is denoted by $\psi = (N_1, N_2, t_1, t_2)$, in which N_1 and N_2 represent the relative effective population sizes of the second and the third epochs, respectively, and t_1 and t_2 indicate the durations of the second and the third epochs, respectively.

We estimate the "neutral" parameters in the model (ψ and θ) from sites putatively free from selection (see Online Methods) by forcing $S = 0$ and maximizing the likelihood of the PRF model. Because there are millions of neutral sites in the Yoruba data set, however, we approximately estimate the neutral parameters in two steps. First, we estimate ψ using only the polymorphic sites in neutral regions (with $1 \leq m_i \leq M_i - 1$). Let $P'(m_i|S = 0, \psi, \theta)$ represent the sampling distribution for these polymorphic sites, with,

$$P'(m_i|S = 0, \psi, \theta) = \frac{P(m_i|S = 0, \psi, \theta)}{\sum_{n=1}^{M_i-1} P(n|S = 0, \psi, \theta)}. \quad (4)$$

ψ is estimated by maximizing a composite likelihood function for these sites alone, ignoring linkage between sites,

$$\hat{\psi} = \arg\max_{\psi} \prod_{i \in \mathcal{S}} P'(m_i|S = 0, \psi, \theta = c), \quad (5)$$

where $\mathcal{S}$ is the set of segregating sites and c is an arbitrary small positive constant. Importantly, this estimator for $\hat{\psi}$ is invariant to the choice of c. This step provides a good approximation under the PRF model (which also ignores linkage), because the probability of the monomorphic sites depends primarily on θ and only very weakly on ψ.

In the second step, we fix $\psi = \hat{\psi}$ and estimate θ using both monomorphic and polymorphic neutral sites. However, to reduce computational cost, we consider only a random sample of 5% of monomorphic and polymorphic sites. Because θ is a single scalar parameter and the data set is large, the downsampling procedure has a negligible impact on the accuracy of the estimated θ. Specifically, we calculate the probability of observing a monomorphic site if both monomorphic and polymorphic sites are included as,

$$P(m_i = 0|S = 0, \hat{\psi}, \theta) = 1 - \sum_{m_i=1}^{M_i-1} P(m_i|S = 0, \hat{\psi}, \theta). \quad (6)$$

Then we estimate θ by maximizing the composite likelihood,

$$\hat{\theta} = \arg\max_{\theta} \prod_{i \in \mathcal{R}} P(m_i | S = 0, \hat{\psi}, \theta), \tag{7}$$

where $\mathcal{R}$ is our subsample of monomorphic and polymorphic "neutral" sites.

Mixture model for inferring representative selection coefficients. Given the estimates of the neutral parameters, $\hat{\theta}$ and $\hat{\psi}$, we fit a three-component mixture model to represent the global distribution of fitness effects based on all polymorphic and monomorphic sites in coding regions. The first component in this model describes neutral evolution and its selection coefficient S_0 is fixed to 0 by definition. The second and the third components represent weak negative selection ($S_1 < 0$) and strong negative selection ($S_2 < S_1$), respectively. Let w_0, w_1, and w_2 represent the probabilities of these three mixture components, respectively. All parameters are estimated by maximum likelihood,

$$(\hat{w}_0, \hat{w}_1, \hat{w}_2, \hat{S}_1, \hat{S}_2) = \arg\max_{w_0, w_1, w_2, S_1, S_2} \prod_i \sum_{n=0}^{2} w_n P(m_i | S_n, \hat{\psi}, \hat{\theta}) \tag{8}$$

subject to the linear constraint $\sum_{n=0}^{2} w_n = 1$ as well as the constraints that $S_0 = 0$, $S_1 < 0$, and $S_2 < S_1$. In this likelihood function, $P(m_i \neq 0 | S_n, \hat{\psi}, \hat{\theta})$ is defined in Equation 3 and describes the probability of observing a polymorphic site, and $P(m_i = 0 | S_n, \hat{\psi}, \hat{\theta}) = 1 - \sum_{m_i=1}^{M_i - 1} P(m_i | S_n, \hat{\psi}, \hat{\theta})$ describes the probability of observing a monomorphic site given a selection coefficient S_n. The estimated representative selection coefficients, $\hat{S}_0 = 0$, $\hat{S}_1$, and $\hat{S}_2$, are then fixed in the mixture density network described below. Finally, the unscaled selection coefficients $\hat{s}_n$ that LASSIE reports are calculated using the equation $\hat{s}_n = \frac{2\hat{\mu}\hat{S}_n}{\hat{\theta}}$, where $\hat{\mu} = 1.26 \times 10^{-8}$ is the estimated human non-CpG mutation rate per site per generation[11].

Mixture density network architecture. In the mixture density network[12] for inferring allele-specific selection coefficients, we assume that the allele-specific weights of selection coefficients can be inferred from the vector of genomic features, $\mathbf{X}^{ijk}$, for a mutation event from ancestral allele j to derived allele k at site i. We denote $\mathbb{P}^{ijk}$ as the vector of allele-specific probabilities of being under neutral evolution, weak selection, or strong selection.

We assume that $\mathbb{P}^{ijk}$ is determined by a mixture density network, which, in our experiments, either includes a single hidden layer or no hidden layers. If the mixture density network includes a hidden layer, the hidden layer is defined by,

$$\mathbf{H}^{ijk} = \text{dropout}(\text{ReLU}(\mathbf{X}^{ijk} \cdot \mathbf{W}_{\text{hidden}})) \tag{9}$$

where $\mathbf{W}_{\text{hidden}}$ denotes the weights and bias terms while **ReLU** and **dropout** denote the rectified linear layer[13] and dropout layer[14], respectively. The rectified linear layer serves as an activation function and the dropout layer serves as a regularizer for preventing overfitting.

At the top layer of the mixture density network, LASSIE first calculates an affine transformation of the hidden features and then employs a softmax function to transform the affine transformation to a normalized probability vector, which defines the allele-specific weights of selection components. This probability vector is defined by,

$$\mathbb{P}^{ijk} = \text{softmax}(\mathbf{H}^{ijk} \cdot \mathbf{W}_{\text{output}}), \tag{10}$$

where $\mathbf{W}_{\text{output}}$ denotes the weights and biases associated with the hidden feature vector $\mathbf{H}^{ijk}$, and **softmax** denotes the softmax layer. If no hidden layer is included, Equation 9 is simply replaced by the identity transformation, $\mathbf{H}^{ijk} \equiv \mathbf{X}^{ijk}$.

Objective function of the mixture density network. To estimate the parameters $\mathbf{W}_{\text{output}}$ and $\mathbf{W}_{\text{hidden}}$ in the mixture density network, we need to design a loss function that captures the discrepancy between the predictions from the mixture density network and the observed polymorphism patterns. We define this loss function as the negative logarithmic value of a likelihood function derived from the PRF model. While uncertainty in ancestral alleles can be considered (see above), here we simply assume that the ancestral allele is known for simplicity and speed. Accordingly, only sites with unambiguous ancestral alleles ($q_i > 0.98$) are used in the training of the mixture density network. The scaled mutation rate per site is fixed at $\hat{\theta}$ and we assume an equal mutation rate to each alternative allele, so the scaled mutation rate for each possible derived allele is equal to $\frac{\theta}{3}$.

Under these assumptions, the probability of observing a mutation from ancestral allele j to derived allele k at a polymorphic site i can be calculated as,

$$\mathcal{L}^i = \frac{1}{3}\sum_{n=0}^{2} \mathbb{P}_n^{ijk} Q(m_i|\hat{S}_n, \hat{\psi}, \hat{\theta}), \tag{11}$$

where $Q(m_i|\hat{S}_n, \hat{\psi}, \hat{\theta})$ is defined in Equation 2 and $\mathbb{P}_n^{ijk}$ is the probability that the mutation belongs to selection component n. The probability of observing a monomorphic site is then equal to one minus the total probability of observing a mutation at this site,

$$\begin{aligned}
\mathcal{L}^i &= 1 - \sum_{k\in\{\text{A,G,C,T}\}\setminus\{j\}} \sum_{n=0}^{2}\sum_{m=1}^{M-1} \frac{1}{3}\mathbb{P}_n^{ijk} Q(m|\hat{S}_n, \hat{\psi}, \hat{\theta}) \\
&= \frac{1}{3}\sum_{k\in\{\text{A,G,C,T}\}\setminus\{j\}} \sum_{n=0}^{2} \mathbb{P}_n^{ijk} - \frac{1}{3}\sum_{k\in\{\text{A,G,C,T}\}\setminus\{j\}} \sum_{n=0}^{2}\sum_{m=1}^{M-1} \mathbb{P}_n^{ijk} Q(m|\hat{S}_n, \hat{\psi}, \hat{\theta}) \\
&= \frac{1}{3}\sum_{k\in\{\text{A,G,C,T}\}\setminus\{j\}} \sum_{n=0}^{2} \mathbb{P}_n^{ijk}[1 - \sum_{m=1}^{M-1} Q(m|\hat{S}_n, \hat{\psi}, \hat{\theta})] \\
&= \frac{1}{3}\sum_{k\in\{\text{A,G,C,T}\}\setminus\{j\}} \sum_{n=0}^{2} \mathbb{P}_n^{ijk} Q(m=0|\hat{S}_n, \hat{\psi}, \hat{\theta}),
\end{aligned} \tag{12}$$

where $Q(m=0|\hat{S}_n, \hat{\psi}, \hat{\theta}) = 1 - \sum_{m=1}^{M-1} Q(m|\hat{S}_n, \hat{\psi}, \hat{\theta})$. It is worth noting that $Q(m|\hat{S}_n, \hat{\psi}, \hat{\theta})$ is not dependent on the parameters in the mixture density network and, therefore, can be precomputed and cached for efficient evaluation of the likelihood function. Assuming independence across sites, the loss function for a mini-batch of sites follows

$$\text{loss} = -\frac{\sum_i \log(\mathcal{L}^i)}{k} \tag{13}$$

in which k is the number of sites in a mini-batch and $\mathcal{L}^i$ is defined in Equation 11 and 12.

References

1. Sawyer, S. A. & Hartl, D. L. Population genetics of polymorphism and divergence. *Genetics* **132**, 1161–1176 (1992).

2. Evans, S. N., Shvets, Y. & Slatkin, M. Non-equilibrium theory of the allele frequency spectrum. *Theoretical Population Biology* **71**, 109 – 119 (2007).

3. Kimura, M. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**, 893–903 (1969).

4. Crank, J. & Nicolson, P. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Mathematical Proceedings of the Cambridge Philosophical Society* **43**, 50–67 (1947).

5. Press, W., Teukolsky, S., Vetterling, W. & Flannery, B. *Numerical Recipes in C* (Cambridge University Press, 1992), 2nd edn.

6. Gronau, I., Arbiza, L., Mohammed, J. & Siepel, A. Inference of natural selection from interspersed genomic elements based on polymorphism and divergence. *Molecular Biology and Evolution* **30**, 1159–1171 (2013).

7. Arbiza, L. *et al.* Genome-wide inference of natural selection on human transcription factor binding sites. *Nature Genetics* **45**, 723–729 (2013).

8. Williamson, S. H. *et al.* Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proceedings of the National Academy of Sciences* **102**, 7882–7887 (2005).

9. Boyko, A. R. *et al.* Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* **4**, e1000083 (2008).

10. Racimo, F. & Schraiber, J. G. Approximation to the distribution of fitness effects across functional categories in human segregating polymorphisms. *PLoS Genet* **10**, e1004697 (2014).

11. Rasmussen, M. D., Hubisz, M. J., Gronau, I. & Siepel, A. Genome-wide inference of ancestral recombination graphs. *PLOS Genetics* **10**, e1004342– (2014).

12. Bishop, C. M. Mixture density networks. Tech. Rep., Aston University (1994).

13. Nair, V. & Hinton, G. E. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, 807–814 (Omnipress, USA, 2010).

14. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* **15**, 1929–1958 (2014).

15. Liu, X., Jian, X. & Eric, B. dbNSFP v2.0: A database of human nonsynonymous snvs and their functional predictions and annotations. *Human Mutation* **34**, E2393–E2402 (2013).

16. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864 (1974).

17. Wong, W. C. *et al.* CHASM and SNVBox: toolkit for detecting biologically important single nucleotide mutations in cancer. *Bioinformatics* **27**, 2147–2148 (2011).

18. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research* **20**, 110–121 (2010).

19. Xiong, H. Y. *et al.* The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).

20. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

21. Findlay, G. M. *et al.* Accurate classification of BRCA1 variants with saturation genome editing. *Nature* (2018).

Supplementary Table 1: Genomic features for LASSIE (hg19 assembly)

Feature group	Feature name	Type	Reference	Note
Variant category	Stop-gain	Binary	ref. 15	Indicate if a mutation results in a gain of stop codon
	Stop-loss	Binary	ref. 15	Indicate if a mutation results in a loss of stop codon
	Missense	Binary	ref. 15	Indicate if a mutation results in a substitution of amino acid
Sequence conservation	SIFT prediction	Binary	ref. 15	Binary prediction of deleteriousness from SIFT
	LRT prediction	Binary	ref. 15	Binary prediction of deleteriousness from LRT
	MA prediction	Binary	ref. 15	Binary prediction of deleteriousness from Mutation Assessor
	PROVEAN prediction	Binary	ref. 15	Binary prediction of deleteriousness from PROVEAN
	SLR score	Binary	ref. 15	Raw SLR score
	SIFT score	Numeric	ref. 15	Raw SIFT score
	LRT omega	Numeric	ref. 15	Raw LRT score
	MA score	Numeric	ref. 15	Raw Mutation Assessor score
	PROVEAN score	Numeric	ref. 15	Raw PROVEAN score
	Grantham score	Numeric	ref. 16	Raw Grantham score
	HMM entropy	Numeric	ref. 17	HMM entropy score from SNVBox
	HMM relative entropy	Numeric	ref. 17	HMM relative entropy score from SNVBox
	dscore	Numeric	ref. 15	Dscore from PolyPhen-2
	Primate phyloP score	Numeric	ref. 18	Primate phyloP conservation score
	Mammalian phyloP score	Numeric	ref. 18	Mammalian phyloP conservation score
	Vertebrate phyloP score	Numeric	ref. 18	Vertebrate phyloP conservation score
Structural information	PredRSAB	Numeric	ref. 17	Probability of the residue being buried
	PredRSAI	Numeric	ref. 17	Probability of the residue being intermediately exposed
	PredRSAE	Numeric	ref. 17	Probability of the residue being exposed
	PredBFactorF	Numeric	ref. 17	Probability that the residue's backbone is flexible
	PredBFactorM	Numeric	ref. 17	Probability that the residue's backbone is intermediately flexible
	PredBFactorS	Numeric	ref. 17	Probability that the residue's backbone is stiff
	PredStabilityH	Numeric	ref. 17	Probability that the residue strongly stabilizes folding
	PredStabilityM	Numeric	ref. 17	Probability that the residue stabilizes folding
	PredStabilityL	Numeric	ref. 17	Probability that the residue destabilizes folding
	PredSSE	Numeric	ref. 17	Probability that the secondary structure of the residue is strand
	PredSSH	Numeric	ref. 17	Probability that the secondary structure of the residue is helix
	PredSSC	Numeric	ref. 17	Probability that the secondary structure of the residue is loop
Regulatory information	SPIDEX	Numeric	ref. 19	SPIDEX Splicing score
	Maximum RNA-seq signal	Numeric	ref. 20	Maximum RNA-seq signal from the Roadmap Epigenomics Project

Supplementary Table 2: Model fitting of the mixture density network. All coding sites on chromosome 1 were used as the held-out test data.

Number of hidden layers	Average loss in the held-out test data
No hidden layer (linear)	0.0321644
One hidden layer (Nonlinear)	0.0322484

Supplementary Table 3: Top 10 most enriched Gene Ontology (molecular function) terms among the 1,118 genes under enhanced selection.

Category	Fold enrichment	p-value	FDR
GABA receptor activity (GO:0016917)	5.93	3.63E-04	3.53E-03
voltage-gated potassium channel activity (GO:0005249)	5.07	5.78E-08	1.53E-06
glutamate receptor activity (GO:0008066)	4.98	4.67E-05	6.18E-04
mRNA binding (GO:0003729)	3.85	2.86E-09	1.32E-07
voltage-gated ion channel activity (GO:0005244)	3.73	1.10E-06	2.04E-05
translation initiation factor activity (GO:0003743)	3.46	3.24E-03	2.40E-02
adenylate cyclase activity (GO:0004016)	2.91	4.94E-04	4.57E-03
translation regulator activity (GO:0045182)	2.88	8.03E-04	7.08E-03
chromatin binding (GO:0003682)	2.87	1.86E-06	3.13E-05
ligand-gated ion channel activity (GO:0015276)	2.85	2.01E-05	2.86E-04

Supplementary Table 4: Top 10 most enriched Reactome pathways among the 1,118 genes under enhanced selection.

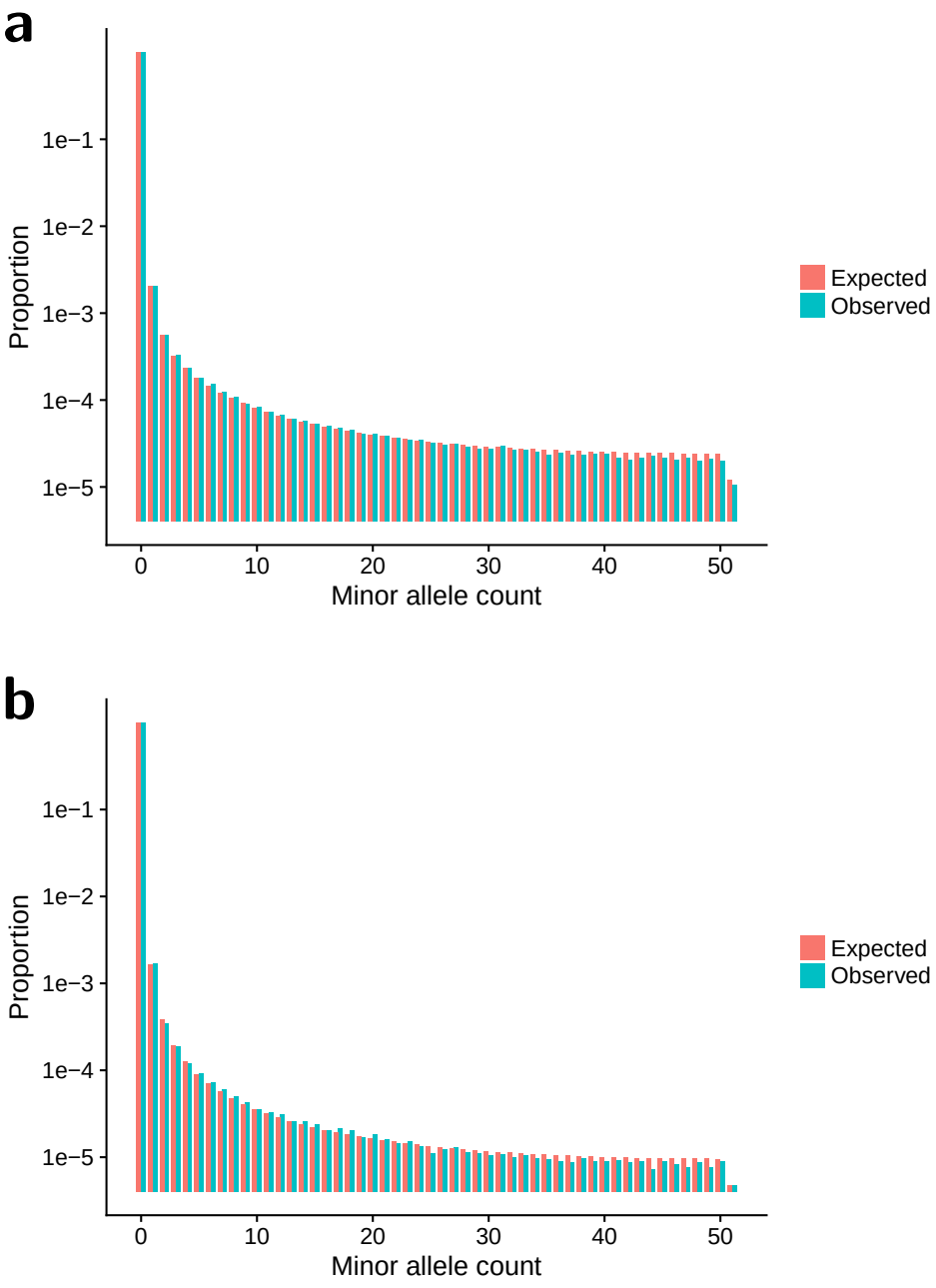
Category	Fold enrichment	p-value	FDR
Cohesin Loading onto Chromatin (R-HSA-2470946)	10.37	7.87E-05	1.04E-03
CREB phosphorylation through the activation of Adenylate Cyclase (R-HSA-442720)	10.37	3.12E-03	2.02E-02
GABA A receptor activation (R-HSA-977441)	9.22	4.11E-05	5.91E-04
PTK6 Regulates RHO GTPases, RAS GTPase and MAP kinases (R-HSA-8849471)	9.08	1.36E-04	1.58E-03
HuR (ELAVL1) binds and stabilizes mRNA (R-HSA-450520)	8.89	4.55E-04	4.46E-03
Adenylate cyclase activating pathway (R-HSA-170660)	8.30	5.22E-03	2.99E-02
Unblocking of NMDA receptor, glutamate binding and activation (R-HSA-438066)	8.15	3.24E-06	6.18E-05
CREB phosphorylation through the activation of CaMKII (R-HSA-442729)	7.98	1.05E-05	1.75E-04
Interleukin-21 signaling (R-HSA-9020958)	7.78	7.37E-04	6.67E-03
PKA-mediated phosphorylation of CREB (R-HSA-111931)	7.54	1.09E-04	1.36E-03

Supplementary Table 5: Enriched Gene Ontology (molecular function) terms among the 773 genes under relaxed selection.

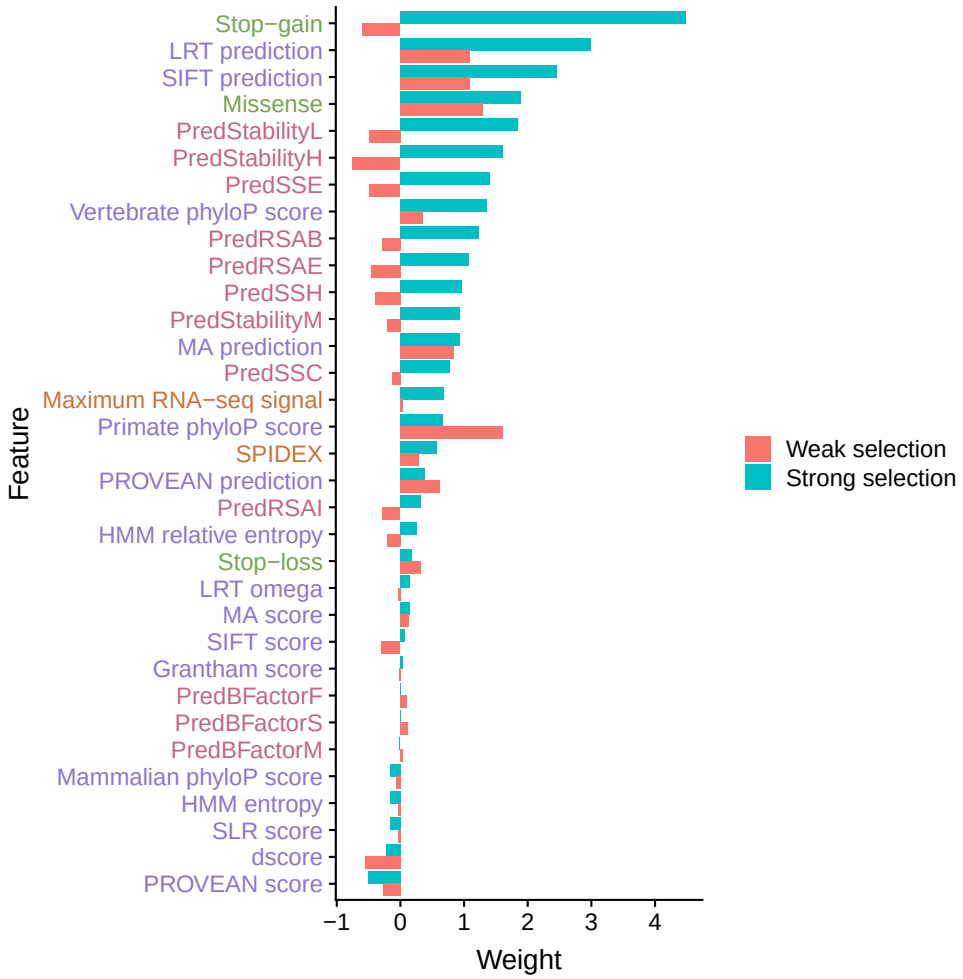
Category	Fold enrichment	p-value	FDR
metallopeptidase activity (GO:0008237)	3.57	2.08E-03	4.81E-02
ATPase activity, coupled to transmembrane movement of substances (GO:0042626)	3.12	1.80E-03	6.67E-02
oxidoreductase activity (GO:0016491)	2.82	1.04E-10	1.93E-08
catalytic activity (GO:0003824)	1.45	1.81E-09	1.67E-07

Supplementary Table 6: Top 10 most enriched Reactome pathways among the 773 genes under relaxed selection.

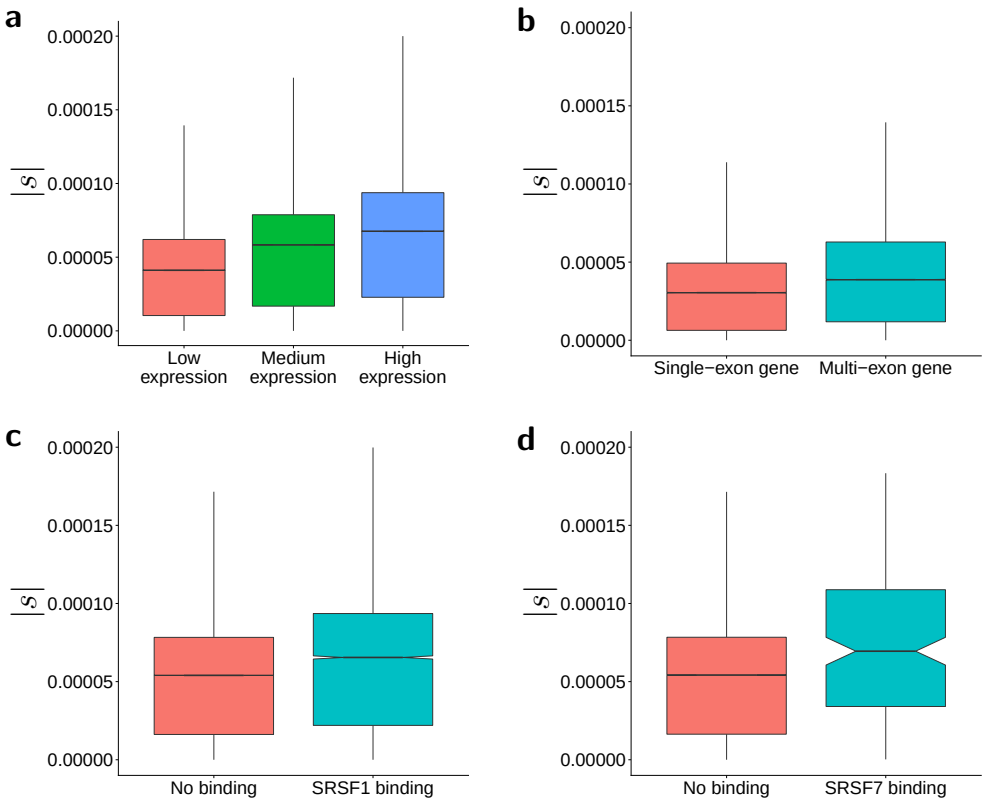
Category	Fold enrichment	p-value	FDR
Melanin biosynthesis (R-HSA-5662702)	15.87	1.42E-04	2.29E-02
Eicosanoids (R-HSA-211979)	12.34	1.79E-05	5.35E-03
Fructose metabolism (R-HSA-5652084)	11.33	4.03E-04	4.45E-02
Laminin interactions (R-HSA-3000157)	6.04	2.07E-04	2.41E-02
Cytochrome P450 - arranged by substrate type (R-HSA-211897)	5.41	1.17E-06	8.20E-04
Phase I - Functionalization of compounds (R-HSA-211945)	4.31	8.54E-07	8.95E-04
Collagen formation (R-HSA-1474290)	3.84	3.82E-05	8.01E-03
Collagen biosynthesis and modifying enzymes (R-HSA-1650814)	3.71	5.19E-04	4.94E-02
Diseases of metabolism (R-HSA-5668914)	3.27	4.77E-04	4.77E-02
Biological oxidations (R-HSA-211859)	2.82	6.63E-06	3.48E-03



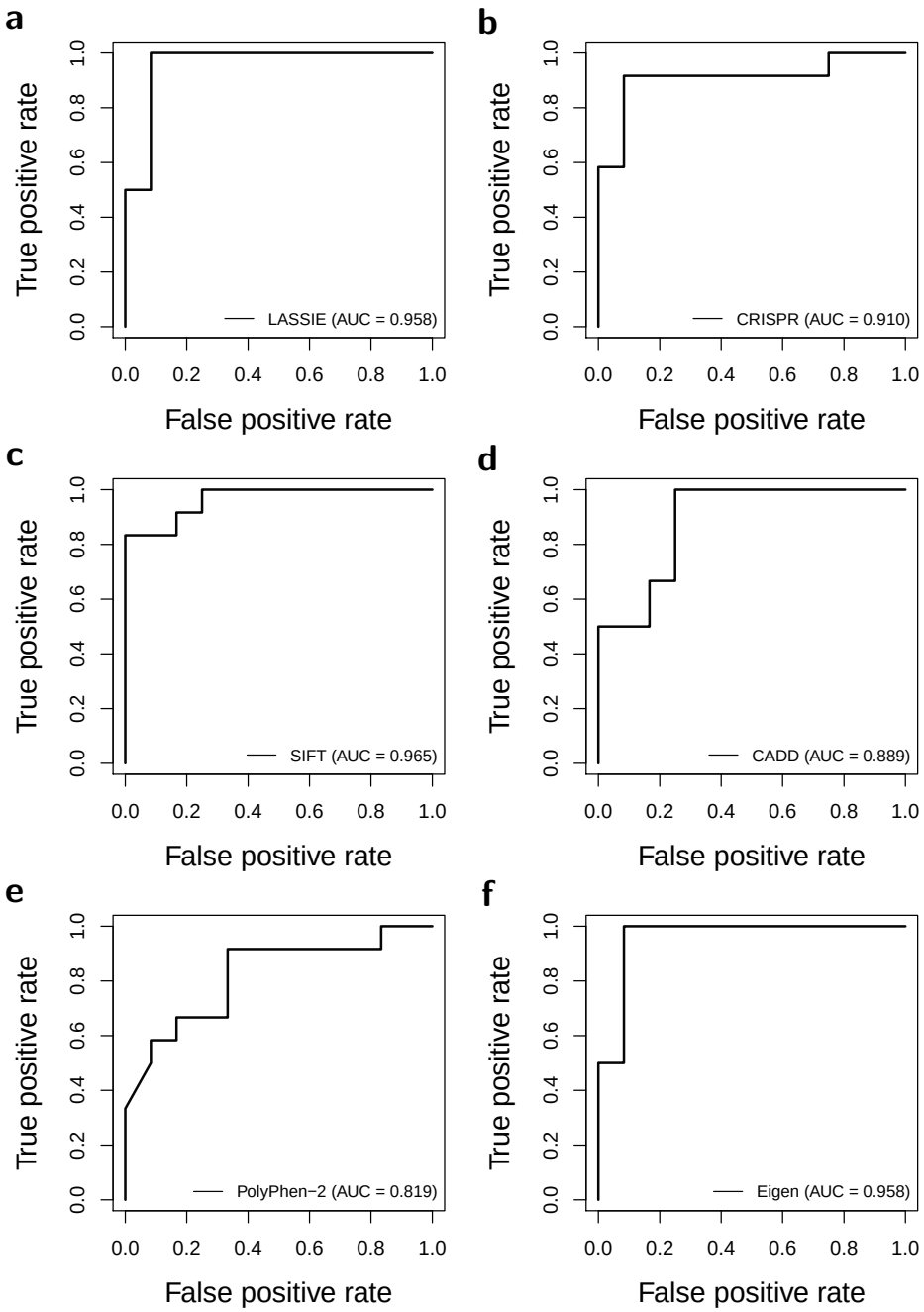
Supplementary Figure 1: Comparison of folded site-frequency spectra between the Poisson Random Field model and the observed data. **(a)** The expected site-frequency spectrum from the demographic model provides an excellent fit to the observed site-frequency spectrum in putative neutral regions. **(b)** The expected site-frequency spectrum from the three-component selection model provides an excellent fit to the observed site-frequency spectrum in coding regions.



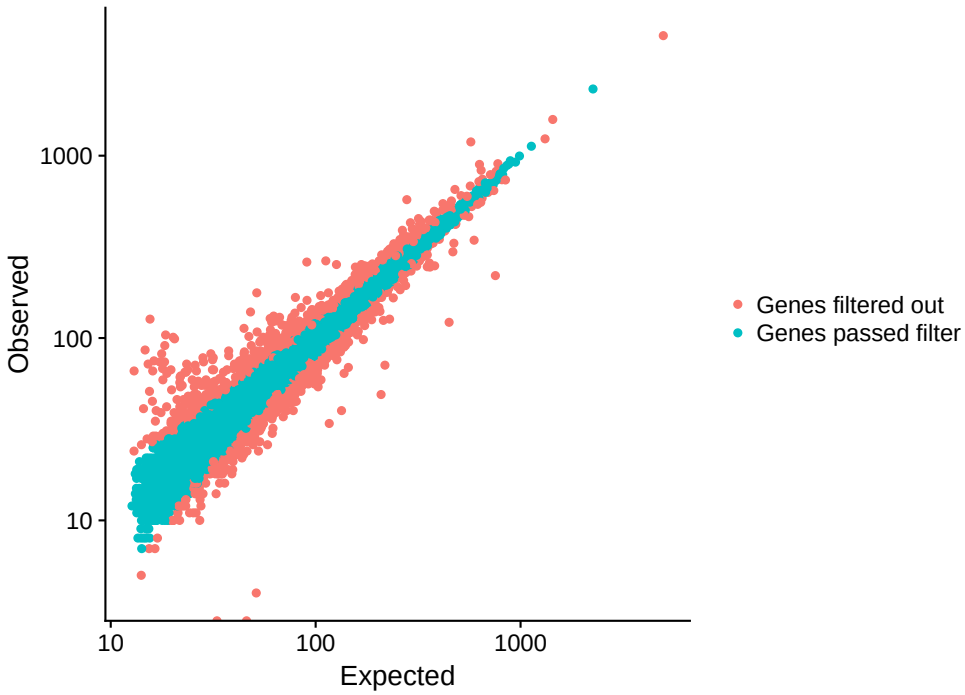
Supplementary Figure 2: Feature weights estimated by LASSIE. Blue and red bars depict the weights associated with strong and weak selection, respectively. A positive weight suggests that the corresponding feature is positively correlated with weak or strong selection. The colors of feature names correspond to four feature groups: variant category (green), sequence conservation (purple), structural information (red), and regulatory information (orange).



Supplementary Figure 3: Distribution of selection coefficients across synonymous mutations. (**a**) Negative selection on synonymous mutations is positively correlated with gene expression level (Spearman's rank correlation coefficient $\rho = 0.286$; two-tailed $p < 10^{-15}$ by t-test). (**b**) Negative selection on synonymous mutations is stronger in multi-exon genes than single-exon genes (two-tailed $p < 10^{-15}$ by Wilcoxon rank-sum test). (**c**) Negative selection on synonymous mutations is stronger in SRSF1 binding sites than non-binding sites (two-tailed $p < 10^{-15}$ by Wilcoxon rank-sum test). (**d**) Negative selection on synonymous mutations is strong in SRSF7 binding sites than non-binding sites (two-tailed $p = 3.036 \times 10^{-10}$ by Wilcoxon rank-sum test). In each box plot, the bottom, the top, and the internal horizontal bar of each box depict the firt quartile, the third quartile, and the median, respectively. The whiskers represent the 1.5-fold interquartile ranges.



Supplementary Figure 4: Prediction power of different computational and experimental methods for separating pathogenic variants from benign variants in the *BRCA1* gene. The CRISPR scores were obtained from Findlay *et al.* 2018 (ref. 21).



Supplementary Figure 5: Comparison of expected and observed numbers of synonymous mutations across all protein coding genes in the ExAC data set. Each dot represents a single protein coding gene. A gene was filtered out if it is enriched or depleted with synonymous mutations in the ExAC data set (Poisson-Binomial test; FDR rate $\leq$ 0.2). Genes with less that 200 potential synonymous mutations were also filtered out and were not shown in the plot.